

ECONOMIC PRINCIPLES AND CONCEPTS AS APPLIED TO MUNICIPAL WATER UTILITIES

FINAL REPORT



GeoEconomics Associates Incorporated

**Jeff Harris and Donald Tate
in association with
Steven Renzetti
and
Acres Associated Environmental Limited**

ACKNOWLEDGEMENTS

This study was conducted for the Ontario Superbuild Corporation (project number SSB-018197.) GeoEconomics Associates Incorporated (GA) would like to acknowledge the work of Dr. Steven Renzetti, particularly for the use of his published econometrics papers in Chapters 3, 4, and 5, but also for his helpful comments. GA would also like to acknowledge the work of Acres Associated Environmental Limited for their work on the engineering aspects of cost in Chapter 5. GA also acknowledges the work of Andréanne Boisvert, Anton van Heusden, and Lisa van Heusden in finalizing the report. Finally, we have benefited from the comments of members of the Ontario SuperBuild Corporation staff, and we have appreciated the high degree of cooperation, collegiality and insight provided.

While acknowledging all of this assistance, we stress that GA is completely responsible for this report, and that any errors, omissions, and interpretations are ours alone.

Donald Tate, President,
GeoEconomics Associates Incorporated

July, 2002

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	3
TABLE OF CONTENTS	4
LIST OF TABLES	6
LIST OF FIGURES	8
ABSTRACT.....	9
EXECUTIVE SUMMARY	10
CHAPTER 1 INTRODUCTION.....	13
1.1 Background	13
1.2 Purpose of the Report	13
1.3 Key Economic Issues	14
1.4 Scope of the Report.....	15
1.5 Main Message of the Report.....	16
1.6 Overview of the Report	17
1.7 Characteristics of the Municipal Water Industry: An Overview	17
CHAPTER 2: KEY ECONOMIC PRINCIPLES AND CONCEPTS	23
2.1 Principles	23
2.2 Basic Economic Concepts.....	26
CHAPTER 3: MUNICIPAL UTILITIES IN A WATER MANAGEMENT CONTEXT	36
3.1 Introduction.....	36
3.2 Economic Theory and Water Management	36
3.3 Water Quantity Issues in Ontario: An Overview	41
3.4 Water Quality Issues in Ontario: An Overview	45
3.5 Summary	56
CHAPTER 4: THE DEMAND FOR MUNICIPAL WATER SERVICES.....	58
4.1 Introduction.....	58
4.2 Municipal Water Demands in Ontario.....	59
4.3 Nature and Characteristics of Municipal Water Demand.....	61
4.4 Dependent Variables: Water Supply and Waste Treatment Demands.....	62
4.5 Independent Variables: Economic Factors	65
4.6 Independent Variables: Structural Factors	69
4.7 Other Independent Variables	73
4.8 The Demand for Residential Water Services – Theoretical Studies	74
4.9 Water Demand Management.....	86
4.10 Summary	89
Appendix 4.1: Price and Income Elasticities	90
CHAPTER 5: WATER SUPPLY AND ASSESSMENTS OF WATER UTILITY COSTS	95
5.1 Engineering Practice in Water Servicing: An Introduction	95
5.2 Cost Estimating Methodology.....	98

5.3	Costing - Part A: Design Standards	101
5.4	Costing Part B: Cost Estimate Development	108
5.5	Econometric Studies of Water Utility Costs	117
5.6	Summary	127
CHAPTER 6: WATER UTILITY PRICING THEORY AND PRACTICES		128
6.1	Introduction.....	128
6.2	What Is Full Cost Recovery?.....	128
6.3	Economic Theory Related to Water Utility Pricing	134
6.4	Practical Methods and Issues in Water Pricing	150
6.5	Water Pricing in Ontario: A Brief Overview	154
6.6	An Overview of Empirical Studies Related to Future Water System Costs in Ontario	158
6.7	Summary	166
Appendix 6.1: Capital Cost Accounting, Utility and Cash Based Accounts		169
Appendix 6.2: The Base-Extra Capacity Method		171
Appendix 6.3: Mathematical Appendix to Chapter 6.....		173
CHAPTER 7: WATER UTILITY ORGANIZATION AND MANAGEMENT.....		175
7.1	Introduction.....	175
7.2	Economic Theory	176
7.3	Empirical Studies Related to Utility Organization.....	181
7.4	Innovative Financing Measures	185
7.5	Summary	187
CHAPTER 8: PRINCIPAL FINDINGS.....		190
8.1	Introduction.....	190
8.2	Key Principles and Economic Concepts.....	190
8.3	Water Utilities in an Water Resource Economics Context.....	191
8.4	Characteristics of Demand for Water Services.....	192
8.5	Characteristics of the Supply of Water Services	193
8.6	Water Utility Pricing Theory and Practice	194
8.7	Economic Theory and Empirical Results Related to Water Utility Organization and Management.....	196
REFERENCES CITED.....		199

LIST OF TABLES

Table 1.1	Selected Water Use and Pricing Characteristics Ontario 1999.....	13
Table 3.1	Selected Water Quantity Characteristics by Hydrologic Region.....	40
Table 3.2	Annual Runoff (m ³ per second) for Major Ontario Hydrologic Regions.....	41
Table 3.3	Historical Growth in Annual Water Use, 1972-1996.....	42
Table 3.4	Great Lakes Public Water Supply by Sub-basin 1991 and 1996.....	42
Table 3.5	Annual Overall Compliance for 1998.....	46
Table 3.6	Summary of Total Costs and Annual O&M Costs of Groundwater Contamination.....	51
Table 3.7	Summary of Estimated Economic Costs of Walkerton.....	52
Table 4.1	Water Use in Ontario Municipalities by User Group and Population Size Group	58
Table 4.2	Ontario Municipal Water Intake in the Context of Total Provincial Water Use 1999.....	59
Table 4.3	Water Use in Metered and Flat Rate Areas in the Western United States.....	68
Table 4.4	The Effects of Metering on Municipal Water Use—Selected Studies.....	69
Table 4.5	Monthly Domestic Water Use per Capita, by Population Size Group.....	69
Table A4.1	Price and Income Elasticities in North America.....	88
Table 5.1	Recommended Contingency Allowance.....	96
Table 5.2	Sizing Criteria for Water Works Infrastructure.....	99
Table 5.3	Sizing Criteria for Wastewater Infrastructure.....	100
Table 5.4	Ambient Water Quality Indicators.....	101
Table 5.5	Typical Wastewater Treatment Requirements.....	102
Table 5.6	Cost Curve Variables for Construction.....	105
Table 5.7	Cost Curve Qualifiers.....	106
Table 5.8	Percentage of Total Facility Cost Based on Historical Data.....	107
Table 5.9	Other Factors Impacting Cost Curves.....	107
Table 5.10	Marginal Cost and Scale Economies for Water Supply and Sewage Treatment.	118
Table 6.1	An Example of Cost Allocation using The Base–Extra Capacity Method.....	144
Table 6.2	Number and Percentage of Volumetric and Flat Rates, by Population Size Group.....	148
Table 6.3	A Summary of Ontario Monthly Water Prices 1991–1999.....	149
Table 6.4	Typical Prices for Popular Liquids.....	150
Table 6.5	Water Servicing Revenue Fund Expenditures (\$million) for Ontario 1997.....	151
Table 6.6	Water Servicing Cash investment (\$million) for Ontario 1997.....	151
Table 6.7	Water System Revenue Sources 1997.....	152
Table 6.8	Summary of Estimated Water System Revenues and Costs (\$ billion, 1993) to Achieve Water Servicing System Adequacy.....	153
Table 6.9	Revenue Impacts (\$ 10 ⁶) of Rate Level Modifications, 1991.....	154
Table 6.10	Capital Requirements (\$ 10 ⁶) for Water Supply and Waste Treatment, 1995-2005.....	155
Table 6.11	Water and Wastewater Revenue Fund Expenditures and Water Rate Revenue	155
Table 6.12	Using Full Cost Rates to Meet Water System Revenue Shortfalls.....	155
Table 6.13	Summary of Investment Requirements (\$ million) by the CWWA, Ontario, 1997 – 2012.....	156
Table 6.14	The Estimated Cost of Additional Water System Expenditure, 1997.....	157
Table 6.15	Summary of Water Servicing Costs (\$ million) from the Powell Study.....	157
Table 6.16	Comparison of <u>Additional</u> Annual Water System Expenditures \$10 ⁶) for Ontario Water Systems.....	158
Table 7.1	French Contractual Arrangements.....	171

LIST OF FIGURES

Figure 1.1	A Simplified Water Servicing System.....	18
Figure 2.1	Optimal Production Under Conditions of Rising Average Costs.....	25
Figure 2.2	Optimal Production Under Conditions of Declining Average Costs.....	27
Figure 2.3	Equilibrium and Pollution.....	31
Figure 4.1	Typical Water Demands Curves Indicating Varying Price Elasticities.....	63
Figure 4.2	An Individual's Demand Curve for Water.....	65
Figure 5.1	Greenfield Unit Construction Cost for Water Treatment Plants.....	105
Figure 6.1	Alternative Concepts of Full Cost Recovery.....	123
Figure 6.2	The Effect of Cross-Subsidization.....	126
Figure 6.3	Pareto Improvements through Declining Block Rates.....	134
Figure 6.4	Peak Load Pricing with Continuous Capacity Peak Load Pricing with Continuous Capacity.....	137
Figure 6.5	Peak Load Pricing and Discontinuous Capacity.....	139
Figure 6.6	Price Movements over Time.....	140

ABSTRACT

This report presents a detailed analysis of the economic aspects of municipal water servicing. The report begins with four contextual chapters, economic principles and concepts, water utilities and water resource management, the demand for water services, and the supply of water service. The two chapters that follow contain the core information for the report, specifically the theory and practice of water pricing, and alternative organization, ownership and management arrangements for municipal water utilities. A detailed summary concludes the report.

The key principles described are economic efficiency, full cost recovery, enhanced market competitiveness, equity, practicality, and environmental sustainability. The key economic concepts described are marginal cost pricing, economies of scale and scope, natural monopoly, public goods, externalities, water as a renewable, but depletable, resource, and property rights.

The relationship between water utilities and water resource management is discussed in terms of the supply and demand for water resources. The chapter initially reviews some of the economic theory related to water management, then describes water quantity and quality issues in Ontario, closing with selected estimates of pollution related costs to water utilities.

The following chapter describes and assesses some key considerations relative to the demand for water services. Initially the chapter describes water demands in Ontario, then describes the nature and characteristics of water demand, including concepts such as price and income elasticities and peaking behaviour. The chapter concludes with some comments on water demand management.

The description of considerations related to the supply of water services is divided between engineering methods for estimating costs, and econometric studies. The engineering section discusses both design standards and cost estimate development. The econometric section reviews studies assessing four principal variables or types of variables, plant capacity, the rate of capacity utilization, spatial variables, and the services provided.

The next chapter reviews the theory and practice of water utility pricing in Ontario. The theoretical section reviews the economic theory relevant to water utility pricing, including a possible definition of full cost recovery. The practical section describes actual rate setting methodologies in use in Ontario, and current price structures and levels, as well as reviews some current studies assessing the extent of full cost recovery.

The final chapter describes some alternative organization, ownership and management arrangements for municipal water utilities. Initially the chapter reviews some relevant economic theory in the area of enhancing competition, transaction costs and incentive regulation. Econometric studies are reviewed based on alternative experiences in France, England and the United States. The chapter concludes by describing some alternative financing mechanisms.

EXECUTIVE SUMMARY

This report contains a detailed examination of the economic factors that underlie the provision of municipal water services, including both water supply and waste treatment. In-depth analyses include: key general economic characteristics, municipal water servicing in the broader context of water resource economics, factors conditioning both water demand and supply, water pricing, the economic characteristics of alternative methods of service provision, and the issue of long term capital financing. The focus of the report is the Province of Ontario, although theoretical and empirical information is drawn from provincial, national, and international sources.

The first chapter emphasizes the possible current underfunding of municipal water utilities, which derive most of their revenues from local sources. Illustrated by a simple table, major features of this funding shortfall include low water prices and ineffectively structured water rates. This lack of efficient funding may be resulting in inadequate funds to cover required future capital and operating expenses. This issue lies at the heart of the report.

The second chapter focuses on two important issues. The first is an outline of a number of principles, various combinations of which underlie the operation of most water utilities. These include: economic efficiency, equity, full cost recovery, practicality, enhanced market competitiveness, and environmental sustainability. The second focus is on explaining general economic concepts that pertain to water utilities, such as marginal cost pricing, scale and scope economies, and several characteristics that make the economics of water utility management a considerable challenge: the nature of externalities, natural monopolies, public goods, property rights, and the nature of renewable but depletable resources. All of these principles and concepts currently underlying municipal water management imply a task that has significant complexities not faced by purely private sector establishments.

Chapter Three examines the municipal water utility industry against a background of water resource economics, to demonstrate that the latter has a large, but possibly understated, role in influencing the former. The chapter proceeds from an overview of two possible overall management approaches to in-stream water: 1) legal, regulatory methods, and 2) a less employed approach based on economic incentive. Then, the predominant issues of water quantity and quality are examined in the Ontario context. In both areas, the approach is to outline some of the principal features of the Ontario situation, to outline some key economic approaches (e.g., pricing, discharge fees, tradable permits, etc.) that have been developed to address quantity and quality issues, and finally to assess briefly the predominant Ontario approaches to water management, and how they influence and condition the economic and financial practices of water utilities. In terms of quantity, Ontario is relatively water rich, while water quality is more problematic, particularly with regard to non-point pollution, for example, from agriculture. Some cost estimates related to the costs of pollution to water utilities are presented, including economic costs of \$ 155 million from Walkerton, and \$ 125 million over ten years for groundwater contamination from chlorinated solvents.

The report then turns to the subject of municipal water demands, which condition the size, complexity, and cost of municipal water utilities. Initially, a conceptual "model" is developed to describe the factors that combine to determine the levels and time patterns of municipal water demands. Emphasis here is placed on economic factors, although social and physical factors are also included. Then, the chapter examines the results of a number of applied econometric studies of municipal water demands, to illustrate more precisely the mathematical characteristics of various types of municipal water demand, especially focussing on price and income elasticities of demand. These studies show that municipal water demands have statistically significant economic dimensions, which could prove useful in lowering

demands in the future, thereby lowering long term capital requirements. The chapter also explores briefly the concept of water demand management in a municipal context.

The fifth chapter examines the economic issues involved in supplying municipal water services. The primary focus here is on the factors determining the costs of servicing, beginning with an overview of costing water systems from an engineering viewpoint. There are various levels of detail used in determining these costs, depending on the level of precision required. The focus here is on the master planning level. A central component of this discussion is on the derivation of cost curves (e.g., total costs as functions of size) for major utility components. The discussion then turns to the issue of how these costs are dealt with in economic terms, based on an outline of a number of econometric studies in this field. These studies indicate generally that, given fixed intake supply and health standards, water utilities exhibit long run increasing returns to scale, long run economies in both capital and operating costs, as well as short run economies in operating costs as capacity utilization increases. Diseconomies of scale are observed in relation to growth of the distribution network as population densities decline on the perimeter of urban areas. In addition to these economies of scale, economies of scope in supply also appear to exist between services to distinct user classes (residential, industrial), while economies of scope between sewage and water supply are not found in the literature, possibly indicating some scope for disaggregation of typical water utilities. An Ontario study estimating water supply and sewer marginal costs found that marginal costs in Ontario far exceed prices charged.

Chapter Six addresses the issue of water utility pricing, aimed at attaining full cost recovery based on marginal cost pricing principles, following the discussions presented in Chapter Two. The criteria suggested for achieving these ends include: peak load pricing; recovery of marginal distribution network costs through connection charges; and forward looking volumetric rates, incorporating long run marginal cost pricing, and estimates of future, rather than sunk, capital costs. A number of means are outlined for overcoming the revenue shortfalls inherent in using marginal cost pricing for enterprises that experience significant economies of scale, including: 1) subsidization by senior levels of government, and pricing at marginal costs; 2) Coase two part tariffs (volumetric pricing at marginal cost levels, combined with additional fixed access fees chargeable to public utility users); 3) Ramsey pricing (a form of price discrimination that charges higher prices to users with inelastic demands higher prices than to those with elastic demands); and 4) Pareto-optimal linear outlay schedules, or volume discounts for large users. The chapter also discusses the theory of peak load pricing. The chapter then contrasts pricing methods based on average cost criteria (as exemplified by the commonly-used pricing method suggested by the American Water Works Association) with those based on marginal cost pricing (as exemplified by the more recent Canadian Water and Wastewater Association rate manual). The chapter also describes current Ontario water rates and reviews current studies assessing the degree of full cost recovery in Ontario water utilities. While Ontario water utilities appear to be covering their current revenue requirements through user charges, most studies conclude they are not investing and charging sufficiently to meet the long term needs for adequate water infrastructure.

The seventh chapter provides a discussion of alternatives to full public provision available for providing municipal water services in the context of the modern economic theory associated with utility management, for example, transactions costs and price incentive regulation. This is achieved by describing a number of empirical studies of the impacts of alternative practices in three nations: the United Kingdom, France, and the U.S. These countries all employ varying degrees of private sector involvement in service provision, over and above the latter's participation in facility construction. These range from full privatization in England, to predominantly contracted management in France, to limited private ownership in the United States. The main conclusions from the various studies reviewed are that: 1) greater private sector participation, in and of itself, does not necessarily produce more efficient utility management, and 2) that local circumstances and the form of regulation of privatized water utilities are important determinants of the success of private sector involvement in water utility management.

Finally, Chapter Eight summarizes the key findings of the report.

CHAPTER 1 INTRODUCTION

1.1 Background

Ontario municipalities, like many throughout Canada and most developed nations, are facing increasing demands on their financial resources. One effect of this increasing demand is a decline in the resources available to support water utilities¹, one of the most expensive components of municipal infrastructure. The overall problem can be summarized succinctly: there is a lack of monetary resources being devoted to the municipal water servicing sector, particularly given increased demands for public funds, the transfer of ownership of capital assets from the Ontario Clean Water Agency (OCWA) to the municipalities, and projected increases in future spending requirements to meet new or updated standards, such as new drinking water guidelines.

Recognizing this complex issue, the Ontario SuperBuild Corporation (OSBC) has commissioned a series of eight major studies to assess the various issues involved in analyzing this problem. This report comprises the second study in this series. It outlines the principal economic dimensions involved in providing municipal water services, by identifying the major economic principles that underlie this task, and by providing an in-depth overview of relevant research findings pertaining to these principles. Such knowledge will help in planning the provision of adequate municipal water servicing in the future.

1.2 Purpose of the Report

The overall purpose of this report is to describe, analyze, and synthesize the economic principles and considerations that pertain to municipal water utilities. These principles will then be applied in the following key areas:

- pricing of water services, with an emphasis on full cost recovery and economic efficiency;
- organization, ownership and management of water utilities; and
- long-term management and capital financing of water utility assets.

The economic issues to be discussed are grouped into five areas:

- the economic nature of water as a renewable resource, focusing on the impact of in-stream water quantity and quality on water utilities, the public health and environmental externalities associated with operation of these utilities, and the public goods nature of water;
- the factors affecting the demand for municipal water services, including price and income elasticities of demand, peaking characteristics, and the willingness to pay for these water services;
- the factors affecting the supply of water services, separated into components (e.g. treatment, distribution, fire protection, etc.), cost elasticities, economies of scale and scope, the nature of sub-additivity and natural monopoly, the impact of differing technologies and regulations; optimal pricing theory, including full cost recovery, metering, alternative rate setting models, equity and efficiency, and the issue of cross-subsidization of costs within and between user groups;
- the pricing of municipal water services
- economic theory regarding water utility organization and management, including innovative financing mechanisms, public/private partnerships and other forms of utility organization.

¹ Water utilities, as used throughout this report, include both water supply and waste treatment function. These may also be referred to as "municipal water servicing systems".

1.3 Key Economic Issues

Many factors are involved in analyzing the economic aspects of municipal water infrastructure; these will be addressed in the appropriate chapters of this report. However, it is useful here to provide a brief overview of a few key issues involved, as a means of introducing the general topic.

Selected dimensions of the current economic problem are illustrated in the following table.

Table 1.1 Selected Water Use and Pricing Characteristics, Ontario, 1999	
(a) Water Servicing (Population served, '000 persons)	
- Water Supply	9,175
- Sewage Collection	8,896
- Sewage Treatment	9,902
(b) Number of Municipalities Surveyed	261
(c) Population in Municipalities that have Partially or Fully Metered Water Supplies	7,647
(d) Water Rate Types (Number of Rate Structures – Residential Connections)	
- Flat Rates	142
- Constant Unit Charges	119
- Declining Block Rates	58
- Increasing Block Rates	23
- Total	342
(e) Average Total Cost of Residential Service at Selected Use Rates (\$)	
- 10 M ³ /month	23.94
- 25 M ³ /month	32.33
- 35 M ³ /month	36.35
(e) Average Total Cost of Commercial Service at Selected Use Rates (\$)	
- 10 M ³ /month	28.74
- 35 M ³ /month	44.80
- 100 M ³ /month	91.00
Source: Analysis of the Environment Canada Municipal Water Pricing Database , carried out by GeoEconomics in support of OSBC study #4, being led by PriceWaterhouseCoopers LLP.	

A number of important implications and questions may be drawn from this table. First, water prices in Ontario are generally very low, and probably form an insignificant portion of the average consumer's budget. The major issue arising here relates to the adequacy of these prices in recovering the full costs of system establishment, operation and maintenance. Given reduced levels of water system subsidization, local charges by municipalities are now used to fund most current expenditures for municipal water servicing. Central questions arising here are: a) whether these expenditures are sufficient to maintain, replace, and upgrade water systems, and b) whether the prices currently being charged are sufficient to accommodate future capacity expansion, and provide for future upgrades related to water quality.

A second key question arising from Table 1.1 is whether the form, or structure, of water pricing used in Ontario is efficient in sending appropriate price signals to consumers regarding their water use. Commodity prices generally reflect current and future resource availability, and impart "signals" as to resource scarcity. This is a critical issue because it relates centrally to the establishment of appropriately sized, cost effective, water utilities. The table illustrates that many Ontario municipalities continue to employ flat rates for water pricing, reflecting a lack of metering, particularly in smaller municipalities. Flat rate pricing provides no incentive for water or capital conservation because, essentially, they imply that water is free². Flat rates lead to substantially over-capitalized water systems.³ Other forms of water pricing are distorted by the prevalence of engineering-based rate

² in the sense that the marginal cost of water is zero. The marginal cost issue is examined in detail in Chapter Two.

³ For a general discussion of rate structure impacts on water demand, see Chapter 4, and also Tate and Lacelle (1995)

setting methods which rely on average cost pricing to set rates. These rate structures are characterized by dividing individual customers' water usage into "blocks", wherein prices are set per unit (e.g., per cubic meter) of usage. The most common form of these pricing structures are declining block rates, which are characterized by falling unit prices as usage moves into progressively higher blocks. This type of rate structuring encourages excessive water demands, which in turn give rise to higher-than-necessary capital costs. These practices also lead to a poor dynamic allocation of capital over time, because temporal peak demands (e.g. summer demands) are not attenuated, since capacity is over-priced when newly built, and under-priced near full capacity. This reflects the cycle of debt used in financing new capacity, which declines over time. These weak pricing signals also generate low incentives for technological and cost efficient innovation.

A third key issue arising from Table 1.1 relates to the adequacy of financial assistance from senior levels of government for both financing and regulating water utilities. Throughout the latter half of the twentieth century, and continuing today, there has been a relentless growth in the demands for public financing in many areas of the economy. Regardless of the merit of expenditures in any particular areas, this trend has meant that lower levels of public capital are now available for water infrastructure financing. This trend was documented well in 1985 by the Federation of Canadian Municipalities, which showed declining shares of public works budgets were being devoted to water systems (FCM, 1985). Recent reports for the Walkerton Inquiry reinforce observations showing that, in Ontario, capital spending by water utilities in Ontario was lower in 1998 and 1999, than in any of the previous eight years (Fortin et al., 2001) The long-term adequacy of this arrangement may be a source of concern, particularly for smaller communities.

Although not represented in Table 1.1, a fourth important issue implied here is whether current forms of water utility organization and management are optimal for financing and managing water utilities. In Ontario, almost all water infrastructure is publicly owned, with operations generally managed either completely by the municipality, or by plant management contracts with OCWA, a provincial crown corporation. A small, but growing trend, is towards private sector involvement in management, similar to the contracts held by the OCWA. A key consideration is the possibility for expanded private sector involvement in terms of investment and operations of water utilities, which could serve as a solution to the problem of inadequate capital financing.

1.4 Scope of the Report

Due to its complexity, the subject addressed in this report requires several different perspectives on water utility management, implying a broad scope for the investigations undertaken.

First, the project requires an integrated view of municipal infrastructure, which takes into account both water supply and waste water treatment⁴. This integration involves five main functions that are described briefly in Section 1.7 below: water supply acquisition and treatment; water distribution; water use (or demand); wastewater collection; and waste water treatment. Each of these functions can be sub-divided further if required. Although many of the economic principles for dealing with these functions are similar, each has special characteristics requiring modifications of basic theory.

Second, the report focuses on conditions as they pertain to Ontario. Certainly, many of the principles to be analyzed are essentially "free" of geographical influence. However, each geographic region also has its own characteristics (e.g., legal and institutional arrangements, accounting standards, etc.) that must be accounted for during the analysis.

⁴ This stands in contrast with the various recent reports of the Walkerton Inquiry, which concentrate on drinking water supply and source protection. The SuperBuild project is substantially broader in scope than that of the Walkerton Inquiry.

Third, economics, by nature, is a policy-oriented discipline. While this study is technical in nature, many of the findings have policy implications. Where these arise, we attempt to point them out without expressing any views as to the direction that should be chosen, apart from those directions conducive to promoting the use of principles such as economic efficiency, equity, and so forth. We also try to interpret the findings of technical studies in terms of their implications for practical action. In no case, though, do we make “recommendations” for public policy, for this is beyond the scope of the study.

Fourth, the bridge between economics and finance is important. For example, selection of an appropriate water pricing method depends, in part, upon economic principles (e.g., economic efficiency), but also depends on financing concepts (e.g., cost allocation and recovery). Accordingly, this project, while focused on economic principles and theories, will also take into account financial concepts and possibilities.

Fifth, while the proposed project focuses on the municipal sector, its findings have implications for management in other sectors, such as industry. These linkages will be drawn and described, if not analyzed fully.

Finally, we will interpret the long-term nature of the study in an economic context, meaning that capital planning and adjustment will be part of the overall study. This has important implications for water rate design and adoption.

1.5 Main Message of the Report

The issues outlined in section 1.3 are complex ones, and have been the subject of much research, deliberation, and practical thought by large numbers of people over a long period of time. It is easy to become lost in details and concepts in carrying out a review and analysis as all-encompassing as this one. As with many authoritative reports, some of the material presented will be accessible only to specialists. Nevertheless, the overall direction and ultimate messages deriving from the report are straightforward.

The starting point is the apparent shortfall of funds to support adequate municipal water servicing in Ontario, whatever the cause of the shortfall. The goal is analyze various economic means by which this basic problem can be overcome. Physical and financial resources have a wide variety of uses. It is an important goal to attempt to employ all available resources in the most effective manner possible. In the case of water utilities, this means providing an adequate level of service, at costs that are economically efficient, affordable, equitable, and sustainable over the long run. This overall goal is important in view of the role of water in assuring good public health, as well as in providing one of the basic bedrocks of a modern economy.

Also, economic principles offer an important perspective in analyzing the overall problem of assuring adequate water utilities. While there are deficiencies in current economic and financial practices in the water utility industry, there is no doubt that, for the most part, this industry has been successful in eradicating many public health problems, as well as supporting Canada’s largest and most diverse regional economy. To continue this pattern of success, the report examines some of the basic economic reforms that seem to be required. Some of the most important of these are (a) a need to recognize that municipal water demands have many inherent characteristics that can be exploited for improved management, (b) the need for substantial reform to current pricing, and (c) the need to at least explore innovative administrative arrangements. Each of these issues is addressed in detail in the report.

The overall message is accordingly a fairly straightforward one. Major economic reforms, if implemented, can substantially alter the incentive mechanisms currently in place, which foster excessive demands

and overcapitalization, towards ones that promote better efficiency in the use of available physical and financial resources. This report will examine the nature of these changes as thoroughly and understandably as possible. It is important to remember this overall “forest” as the individual “trees” of the argument are presented.

1.6 Overview of the Report

The remainder of this chapter provides a basic description of an idealized municipal water utility, to convey that the subject of this report is very “real,” even though the economic theory contained in this report may, at times, be quite abstract. Chapter Two outlines the basic economic principles and concepts that will serve as theoretical background for much of the assessment in the following chapters. Chapter Three then places water utility management into a resource economics context by documenting both the basic economic theory underlying this task, and the current practice in Ontario regarding the management of water resources pertinent to municipal water utilities. This chapter outlines some of the basic characteristics of water quantity and quality as they affect municipalities, describe briefly the main management approaches being taken currently, and assess the latter in terms of economic principles. Chapters Four and Five respectively address the demand and supply characteristics of water utilities. Here, the aim is to examine issues that should be taken into account in financial and economic decision-making for water utilities. Chapter Six examines the theory and practice of water pricing for water servicing, both in general and in the Ontario context. This will include a discussion of the concept of full cost recovery, in addition to some empirical studies of the possible shortfalls in revenue requirements to assure future financial viability. Chapter Seven then outlines the theory and practice of water utility management and financing, illustrated by international case examples. Finally, Chapter Eight presents the main conclusions of the study.

1.7 Characteristics of the Municipal Water Industry: An Overview

1.7.1 Economic and Public Health Characteristics.

The municipal water industry, although quite different from private firms making up what is normally thought of as a nation’s industrial base, shares many characteristics with these firms. It has outputs consisting of potable water and the treatment of wastes to various levels. It has revenues in the multi-billion dollar range, primarily from user charges for its products. It has many inputs, ranging from raw water itself to various material and energy products. What primarily distinguishes the water industry from other industries is its capital intensity, its characteristics as a local natural monopoly, and its importance in terms of public health.

Water servicing is the most capital intensive by far of all public and quasi-public services (Beecher et al., 1992). Beecher et al. suggested that one simple method of measuring the capital intensity of public services was by using a “capital: revenue” ratio, which divides total capital worth of a particular public service by annual revenue. Typical capital:revenue ratios for airlines, telecommunications, and other “quasi public” services (i.e., where there was significant private sector involvement) fell in the range of 1:1 to 2.5:1 in the U.S. at the time the research was conducted. In contrast, water service system ratios ranged between 5:1 and 16:1, with the lower part of the range pertaining to systems where there was

substantial private sector involvement.^{5, 6} This finding demonstrates, in a simple and abstracted manner, the nature of the capital intensity of the municipal water “industry”, where most of the capital intensity derives from the requirement to provide an extensive physical distribution network both for water supply and sewage, as well as fire protection. The need for a physical distribution network means that water utilities are a type of industry known as a network utility, of which other examples are road, railway, electric, gas and telecommunications utilities.

Water utilities are natural monopolies because they fall into the group of industries in which the basic investment costs are so expensive that it does not pay for more than one firm to supply the product or service. It is much cheaper for the existing enterprise to increase its output or lower its prices slightly than it is for a second firm to make the capital investment necessary to compete. Natural monopolies tend to be regulated in order to protect the public from abuses that may arise from monopoly situations, like restrictions of service or excessive prices. In Canada, this tendency for regulation has lessened somewhat since the 1980s. The local natural monopoly characteristics of water utilities stem not only from their high capital costs, but also from their role in providing organized distribution and treatment of water services, primarily to urban populations. Given the capital intensity of water distribution systems within urban areas, the natural monopoly characteristics of water utilities stem from the cost effectiveness in having only one distribution network, as compared to competing networks. As the density of urban populations diminishes at the perimeter of urban areas and into rural areas, it becomes more cost effective to supply individual households and industry from private wells, rather than from utilities, thus generating the natural limit to the size of the local utility. Therefore, water utilities have the characteristic of being a diverse collection of local monopolies.

A final important characteristic of water utilities is their role in terms of public health. While other utility failures such as gas and electricity may have public health implications, service failure for water utilities can be much more serious for many reasons – the sometimes-rapid onset of illness, the often undetected nature of problems until crisis conditions are reached, and the often quickly fatal nature of resulting illnesses.

All of these characteristics highlight the importance of adequate regulatory controls over water utilities, both within municipal utilities and by senior levels of government.

1.7.2 An Overview of a typical Water Utility

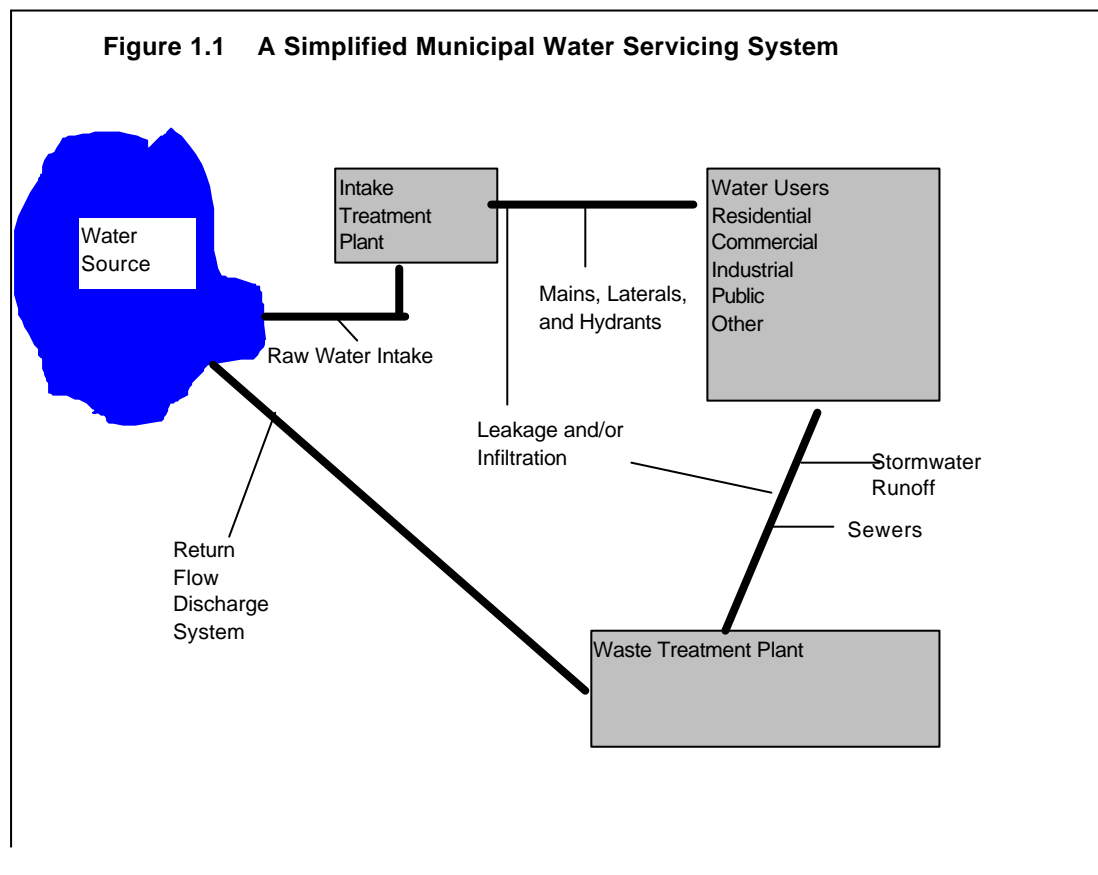
In setting out to describe the economics of municipal water utilities, it is necessary to provide an overview of a typical water utility to provide a context to what could otherwise be an abstract discussion. The purpose here is to identify the various system components to support the discussions of the following chapters.

Figure 1.1 presents a simplified view of a municipal water and wastewater system. Municipalities withdraw water from a **source**, such as a river, lake, or groundwater aquifer. This supplies **intake** to the water treatment plant, where the water is processed to meet potable standards. The treated water forms the **water supply** for the municipality. This water supply is then distributed to satisfy various **water demands**, such as residential, commercial, industrial and others. These demands are the focus of municipal water demand management. After satisfying various demands, water flows via the **sewer system** to a **waste treatment plant**, where it is processed to various levels before being **discharged**

⁵ About 20% of public water servicing in the U.S. was provided through “Public-Private Partnerships” in 1993, a proportion that has probably grown.

⁶ Tate and Lacelle (1995) found that the ratio for Canadian water systems in the mid-1990s was 33:1. These figures denote the very high capital intensity of the industry as a whole, and even hint that there may have been an inefficient use of capital resources in Canada in the past.

back, usually, to its original source. Waste treatment ranges from no treatment at all to tertiary treatment, which removes up to 99% of polluting materials. At any point in the system, water may **leak** from or **infiltrate** the system through pipes that are not water tight, for a variety of reasons. Most municipal water systems normally leak 10%-15% of the water supplied into the system, although percentages greater than 40% have been reported⁷. Paradoxically, high leakage rates may make this the largest demand on a municipality's water supply. Likewise, infiltration into the sewer system may often be so high that more water enters the waste treatment plan than is supplied into the water system. Infiltration may also occur into water supply networks. These types of infiltration can endanger public health if the infiltrating water is contaminated. Both leakage and infiltration pose significant costs on system operation and maintenance, and can be referred to as deadweight losses on the water system. These various components are described in more detail in the next section.



1.7.3 Major Components of a Water Supply System

With regard to **source of supply**, water for municipal purposes in most nations is available for the taking. Sometimes, water abstraction requires permits from provincial, state, or national governments, but these are most often issued without royalties or other fees⁸. Development costs involve largely the costs of constructing, operating, maintaining, and financing the works themselves. In areas where water

⁷In third world nations leakage rates over 60% of total supply have been reported in third world nations (WSSCC, 1998)

⁸ This stands in contrast more business-oriented activities, such as industrial operations, agriculture and power generation. These often are required to pay license fees to the appropriate senior level of government.

allocation occurs on a priority basis, water for municipal purpose is normally assigned the highest priority in the hierarchy of demands. Most municipalities obtain their water from lakes or rivers, with intake systems that can be expanded in stages as required. The remaining systems demand groundwater supplies, which again can often be expanded by drilling more wells. Often the lake or river is sufficiently large that its size does not limit the capacity of the water system. Groundwater supplies are more limited, resulting in a higher need for conservation of water demand, on a total annual basis. Normally the design of the intake or the well capacity will be based on the maximum day requirement of the system.

Water treatment plants function largely to bring raw water quality to potable standards. Their design, accordingly, is a partial function of raw water quality, and the size of the population served. Constructions of water treatment plants occur in stages or modules, being expanded as required. Design is based on the maximum day demand and is largely related to volume (as determined by residential and commercial/industrial demands), with minor portions being determined by fixed costs, (e.g., administration buildings) and fire protection or maximum hour demands.

Pumping facilities both at the treatment plant and in the distribution system, are expandable in stages. Depending on storage in the system, their design is based on flow requirements related to maximum day, maximum hour, and fire protection demands.

Trunk water mains and local distribution networks are usually designed and installed to be adequate for long periods into the future, but may be duplicated if substantial system growth occurs. Their design relates to maximum day flows, maximum hour flows, future growth, and fire protection demands. Major water lines and trunk mains distribute water from the treatment plant to the local networks, and are installed to adequately meet demand for extremely long periods into the future. In some older Canadian cities (e.g., Montreal) some trunk mains are over a hundred years old, and may date from the original construction of the system. With increasing demands these water lines and trunk mains may be duplicated in order to meet current demands.

The local distribution network can be divided into subdivision mains and extensions, street mains, service laterals, meters and hydrants. Storage reservoirs of treated water are usually installed at one or more points in the local system, and will be designed to even out the flow during the day and to provide a reserve for fire fighting purposes, or system operating problems.

Subdivision mains and extensions are sized for the ultimate anticipated development of an area and are rarely duplicated. They will be designed for the maximum hourly demand and fire protection requirements, depending on the local storage capacity installed in the system. Service laterals distribute the water from the subdivision extensions to the individual connections. They are generally sized to meet instantaneous demand, and are rarely duplicated once installed.

Storage of treated water is usually installed at one or more points on the distribution system using either elevated tanks or ground level reservoirs. Storage facilities are designed to attenuate flow variations during the day, provide reserves for both fire fighting purposes and system operating emergencies. Storage reservoirs can be expanded in stages as required.

Distribution mains are sized for the ultimate anticipated development of an area and are rarely duplicated. They are designed to meet maximum hour demands and fire protection requirements. They are usually paid for by the developer of an area and so may not exert a capital cost on the utility.

Service connections are installed to serve the anticipated instantaneous demand, and are usually paid for by the customer without cost to the utility.

Meters are the principal means of measuring water demand by individual connections to the water supply systems. They are normally paid for by individual customers to measure the flow to a particular

connection and are then maintained and calibrated by the utility company. The size of the meter is generally standardized for residential consumers, and requested by industrial or commercial users based on an estimate of maximum daily demand. Meters play a major role in the pricing of water because they not only measure water demand, but also, because their sizes vary, may also be used to approximate peak demands that any given customer can place on system capacity. Meters are also normally used to approximate wastewater flows, as it is rare to measure return flows to waste collection systems. Wastewater flows are often used to divide system revenues costs between the water and the wastewater parts of water utilities, where these form separate agencies within a municipality.

Fire hydrants are usually installed, at their own expense, by developers when development occurs within the municipality. Because they are related almost entirely to fire fighting requirements (but may be used by the utility for flushing purposes), the maintenance and renewal of hydrants are carried out by the utility itself. The costs related to hydrants are charged to the municipal tax base.

1.7.4 Major Components of a Wastewater System.

Service connections to wastewater collection networks on both private and public property are installed to serve anticipated maximum instantaneous flows and are usually paid for by the customer.

Local sewers are installed to meet the ultimate requirements of an area once fully developed. Their design is based on maximum hour flows. Usually they are paid for by the developer of the area, or charged as a local improvement by some municipalities.

Trunk sewers are designed and installed to meet the ultimate flow from a catchment area, which may be larger than the area being developed. Allowing for some flow diversity, their design is based on a peaking factor between maximum day and maximum hour flow, and the total cost will be divided between the several tributary areas within the total catchment area.

Pumping stations are designed structurally to handle the maximum anticipated flow from the tributary area, although all the pumps may not be installed initially. The maximum flow consists of the average annual flow multiplied by peaking factor, somewhere between maximum hour and maximum day.

Wastewater treatment plants are designed hydraulically, usually based on the average annual flow with a peaking factor to take into account the size of the tributary area. Other design parameters may include local experience with stormwater infiltration and inflows into sanitary sewers. Design will largely be based on flow, but some components, (e.g., sludge pumping secondary treatment and nutrient removal facilities) will also be based on the sewage strength. A large portion of the cost of the plant will be for sludge pumping and treatment, which is based on both flow and strength. Wastewater treatment plant components can be built in stages.

Effluent disposal facilities, whether to a water course or to land, may be built in stages. These facilities are sized to adapt to maximum day flows, and their design may incorporate features that reduce the environmental impact on the receiving stream. Sludge disposal can similarly be built in stages and will be dependent on flow, strength and the degree of treatment provided.

CHAPTER 2: KEY ECONOMIC PRINCIPLES AND CONCEPTS

As outlined in chapter 1, municipal water servicing in Ontario has been and continues to be largely a public undertaking. This reflects the current situation throughout Canada, which, in turn, arose largely in recognition of the “natural monopoly” characteristics of this industry. In establishing many of the economic and financial practices of water utilities, such as pricing, a number of principles have developed to compensate for the fact that these utilities are not typical, competitive, market-oriented enterprises.

As used in this report, the term “principles” refers to a set of criteria that can guide both public policy makers and utility operators in making decisions pertaining to water management and water utility operations. The principles outlined in this section are used throughout the report to assess the economic aspects of water management and water utility practices in Ontario, and to discuss possible modifications to those practices. They will also be developed and discussed in more detail in later chapters.

The term “economic concepts” as used here refer to ideas essential to an economic perspective on water resource and water utility management – marginal cost pricing, economies of scale and scope, natural monopoly, public goods and externalities. These will be introduced briefly in this chapter, and used in the following chapters to analyze the economic aspects of municipal water utilities.

2.1 Principles

Economic Efficiency

The basic principle underlying most of economics is that of maximizing social welfare, which is achieved through economic efficiency. Because social welfare is difficult to define empirically, primarily due to difficulties in valuing equity ⁹, economists have developed the more pragmatic concept of Pareto optimality to assess whether society is operating at a social welfare maximum. A Pareto optimum refers to a resource allocation in which no one in society can be made better off by changing that allocation without making someone else worse off. The Pareto optimal condition at which a point of maximum social welfare occurs is defined as a set of efficient conditions associated with the equalization of marginal costs and benefits. The term “marginal” refers to the price or cost of the last output produced by a given industry, or extracted from a natural asset. The two most relevant efficiency concepts pertaining to water utilities occur where:

- marginal cost = price, (the marginal cost pricing “rule”), and
- social costs = social benefits.

The efficiency of these conditions can be illustrated fairly simply by using the example of marginal cost pricing. If marginal cost is less than the price (or willingness to pay of an individual), this means that someone can be made better off by producing more of the good. On the other hand if the marginal cost is higher than the price, no one will buy it and the resources put into its production will be wasted. Thus, only at the equilibrium where price equals marginal cost, does a Pareto optimum occur. The socially efficient marginal conditions refer to the externalities (e.g., public health) and public good nature of water, where a decision made on purely private grounds may not lead to a social optimum, and government intervention is required to achieve social efficiency.

⁹ The principle of equity will be discussed below.

While this discussion has addressed efficiency in a static context, a more relevant context is that of a dynamic economy based on innovation, where efficient prices act as “signals” of current or emerging resource scarcities, and thus indicators of where innovation and/or behavioural change is required. The observation that market prices act as information (Hayek, 1945), or signals to consumers about relative resource values is also related to marginal cost pricing, because the further prices are separated from costs, the more misleading are the signals given, and the more inefficient behaviour becomes. In the case of water utilities, this inefficiency typically results in excess water use, and thus excess capacity and unnecessarily costly infrastructure.

2.1.2 Full Cost Recovery

The principle of economic efficiency is closely aligned with a second basic principle, that of full cost recovery. If full costs are not recovered through water prices, incorrect pricing signals are being given, in this case that water services are cheaper than they actually are. In consequence, consumers are unaware of the real costs of their use of the resource. While seemingly simple, the definition of full cost recovery is actually quite complex, for two principal reasons:

- a) lack of clarity within both economics and accounting as to what constitutes technical capital costs. For example, depending on whether cost- or cash-based accounting methods are used, capital costs will differ. Also, in cost-based accounting, costs will vary depending on the depreciation method used. In economics, replacement cost is not well defined, while substantial controversy is also associated with the definition of the long run marginal cost of capital.
- b) the issue of whether or not to include the social costs associated with environmental degradation, and/or the environment subsidy associated with the use of the environment as a sink for effluent discharge.

2.1.3 Enhanced Market Competition

Another principle related to efficiency is that of enhanced market competition, which applies primarily to the discussion of water utility organization and management, addressed in Chapter Seven. Within economics, it has become broadly accepted that the primary organizational means of securing more efficient utility operation is through restructuring associated with introducing or enhancing competition, as compared to mere changes in the ownership structure. An example would be competitive bidding by private sector firms to operate a municipal water utility. As in the general economy, enhanced competition in the water servicing sector may be expected to generate greater innovation, and possible long run cost reductions, as well as off-loading costs from senior or municipal levels of levels of government.

2.1.4 Equity

A fourth principle used in public water management is that of equity, or what society perceives to be fair. Equity is a difficult concept to deal with in economic terms, because it involves social factors, such as politics and ethics. Economics does not make judgments about whether different social values and behaviours regarding equity are valid or not, but merely describes the consequences of these value judgments in efficiency terms, and observes what those value judgments are. In the context of water utilities, some current equity related social values in Ontario can readily be observed, such as:

- reliable, potable piped running water is perceived as a basic right or entitlement to a greater extent than a normal good, due, in part, to its role in promoting and sustaining acceptable public health;
- society appears to feel an obligation to financially assist small and rural communities with their water infrastructure, both to achieve good public health, and to offset the economic burden of establishing effective water systems;
- large water using industries should be cross-subsidized by residential consumers, and taxpayers for their water use.

While the last observation appears to be rather aberrant, and inconsistent with other social values, it certainly (and consciously so [see AWWA, 1983]) forms the basis for the prevalent average-cost-based, engineering-devised rate setting methods used in Ontario, and throughout North America. This somewhat startling observation will be further explained by in the description of two most common formal rate setting methods recommended by the American Water Works Association (AWWA) – the base-extra capacity and commodity-demand methods in Chapter Six.

2.1.5 Practicality

A fifth basic principle is that of practicality, or administrative ease of use. A key question here is whether relatively sophisticated pricing and utility management methods can be reasonably applied in a large number of local water utilities. While practicality seems to be a reasonable principle, it has often been abused, notably in the common (and overly simplistic) characterization of marginal cost based pricing methods as “impractical” (see for example Fortin et al., 2001), and in discouraging the use of water metering in many communities. In Chapter Six, an attempt is made to “rehabilitate” the idea of marginal cost pricing as a practical, useful concept.

2.1.6 Environmental Sustainability

A final and somewhat newer principle is that of environmental sustainability. This principle is reflected in water management by, for example, appeals for increased water conservation. While both difficult and contentious to define in operational terms, the concept of environmental sustainability can be seen as a non-declining level of environmental or public health over time. Water related examples include: the availability of adequate supplies of potable water; the achievement of non-harmful levels of effluent quality; and, in many countries, the lowering of per capita water use. Assessment and incorporation of this principle in terms of water conservation is straightforward, through the inspection of rate structures. In general, flat rate and declining block rate structures discourage water conservation, while constant unit charges and increasing block rates encourage it.¹⁰ Similarly, for the public health and environmental attributes, stricter regulation or improved environmental pricing, and enforcement, can be expected to improve sustainability, and vice versa.

2.1.7 Conflicting Principles

Each of these principles may be deemed valid in terms of water utility management in particular areas or situations, but often conflict with one another. To use the example of water rates, an administratively simple system (e.g., flat rates) will often be inefficient economically, as well as inequitable and unsustainable. One of the tasks undertaken in this report is the analysis of these various, commonly-used principles in the context of improving the economic practices used in water utility management, to

¹⁰ See Tate and Lacelle (1995) for further discussion.

identify the most important conflicts between these principles, and to draw some conclusions as to which principles are most important in the Ontario context.

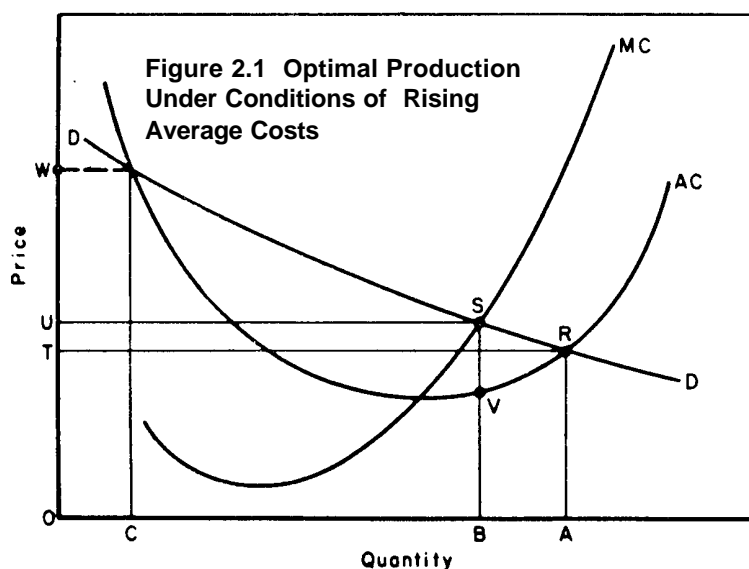
2.2 Basic Economic Concepts

2.2.1 Marginal Cost Pricing¹¹

A fundamental principle of economics, frequently discussed in the context of municipal utility financing and price setting, is that of marginal cost pricing. It is appropriate, therefore, to discuss this concept early in the report.

To do this effectively requires a brief overview of the way in which the analysis of pricing is approached in economics. The mechanics of price determination are generally analyzed in the context of perfectly competitive markets. Although the restrictive assumptions of the perfectly competitive market model are seldom met completely, the competitive market model is surprisingly successful in explaining many features of how goods and services are exchanged in actual markets. While there are few operating, competitive markets in the water resource field¹², the traditional economic model does provide some useful insights into the economic and financial problems faced by municipal utilities. For this reason the following theoretical exposition is worthwhile as background information both here and in the chapter on water pricing practices.

Figure 2.1 illustrates the pricing problem in the context of markets for most goods and services¹³.



¹¹ This section draws heavily on Hirschleiffer et al. (1960), for we consider this to be an effective exposition of applying marginal cost pricing to municipal water utilities.

¹² Water rights trading in the southwestern U.S. is one example of a quasi-competitive water market.

¹³ A working knowledge of the basic principles of microeconomics is assumed here.

In this figure, DD is the demand curve of consumers for the product of a given enterprise. It is assumed that only one class of customer exists, and for this reason a unique average cost function, AC, can be defined as a function of the quantity of the product provided.¹⁴ The curve showing marginal costs as a function of quantity is labelled MC. The latter is necessarily below average cost where the latter are declining, and greater than average cost where the latter are rising, and it follows that the MC curve cuts through the lowest point of the AC curve. If a single price is charged so as to “cover” cost while clearing the market¹⁵, that price can only be equal to OT, since at a price OT the quantity OA would be demanded, the production of which involves an average cost of AR (= OT) per unit.

At this price, zero profits are being earned in an economic sense; price equals unit cost, including a normal interest return on capital invested. However, this solution does not correspond to the best use of society’s resources. This can be shown by considering the production of the units between OA and OB. For each of these units, the marginal cost – the additional cost of supplying the unit considered – is greater than the amount anyone is willing to pay for it – the consumer’s marginal value in use, as indicated by the demand curve. The quantity OB is demanded at the price OU (= BS), and, if any larger quantity is to be taken by consumers, the price will have to be reduced below BS. But the marginal cost is higher than BS throughout the range being considered, which means that there are alternative uses of the resources entering into this marginal cost that consumers value more highly than they value what those resources can produce in the production of water services. The solution for best use of resources is to produce just up to the point where the marginal cost begins to exceed the price that consumers are willing to pay for the additional unit produced; that is to say, the correct output is OB at the marginal-cost price BS.

It can also be noted that the price BS is greater than the average cost BV corresponding to the output OB, so that there is a profit to the private or public enterprise being considered here. Given the existence of profits more firms will enter the market over time, until finally there will be no profit at equilibrium, and marginal costs will equal prices, the result of a market characterized by perfect competition.

2.2.2 Economies of Scale and Scope, and the Concept of Natural Monopoly

a. Economies of Scale

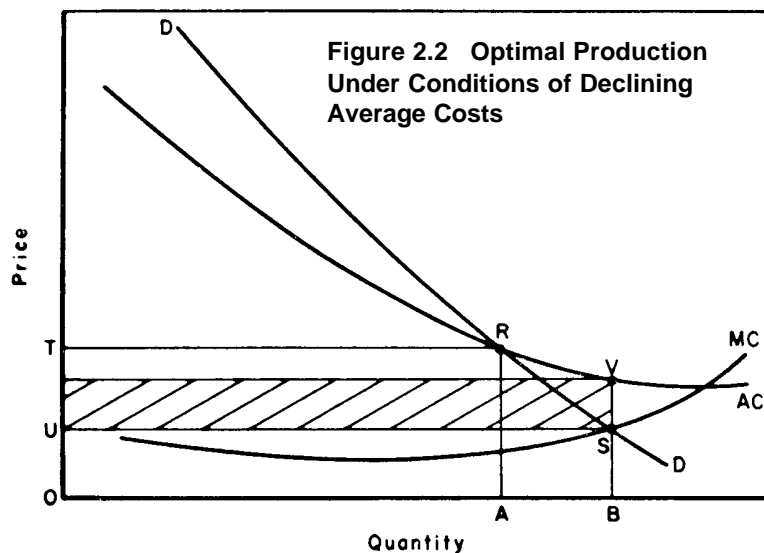
In contrast to perfectly competitive markets, a long standing and pervasive problem in resource economics is that of an enterprise which operates where the demand curve intersects the marginal cost curve at a point where marginal cost lies below average cost. This condition is referred to as economies of scale, a term which refers to the ability to produce an additional unit of production at a lower average cost than the unit preceding it. In the previous section the firm exhibited diseconomies of scale, or increasing average costs at the point of optimal production. For most of this century economies of scale were considered the prime characteristic of natural monopolies, where, given economies of scale, one firm could produce all required production at a lower price than two competing firms.

A problem with marginal cost pricing and efficiency in the presence of economies of scale is illustrated in Figure 2.2. Here the demand curve DD intersects the average-cost curve AC in the range where the latter is still declining. The average-cost output and price are OA and AR, respectively, and the marginal-cost output and price are OB and BS, respectively. But note that in this case the

¹⁴ It is assumed here that the quantity of service includes a bundle of functions consisting of both water supply and waste treatment services.

¹⁵ “Market clearing” refers to the condition whereby all of a commodity produced is consumed, or demanded, by users of the commodity.

marginal-cost output (OB) is greater than OA, whereas in the case considered in Figure 2.1, it was smaller and, correspondingly, the marginal-cost price is here lower than average cost, whereas before it was higher. In consequence, whereas in the previous case the enterprise earned a profit at the marginal cost output and price, here it will incur a loss. The loss (the shaded area in the figure) will be equal to the difference between average cost and price, SV, multiplied by the number of units produced (OB).¹⁶



How this loss is to be made up is essentially a distributional question and the same argument applies for the superiority of the marginal cost price and output over the zero-profit price and output, as outlined in the previous section.

It is clear, however, that the enterprise or the operation illustrated in Figure 2.2 should not be simply abandoned. Evidently, at outputs less than OA the price consumers are willing to pay exceeds average cost, so that a profit *can* be made. At such outputs consumers prefer resources to be invested in this industry rather than elsewhere, so the enterprise should certainly produce up to OA. But, along the lines of the efficiency argument made earlier, the output OB is clearly the best on efficiency grounds, and therefore preferable to outputs less than OA which we know to be already desirable. Should the demand curve DD lie *entirely* below the average-cost curve, then there would be no point at which a profit would be possible so long as a single price was charged. In such circumstances the enterprise may or may not be desirable on efficiency grounds; the fundamental criterion is whether or not the *aggregate* value in

¹⁶ An *economic* loss, which is what is being discussed here, is not necessarily the same as an *accounting* loss. The differences between the two are generally discussed in elementary economics textbooks. For present purposes the most important difference is that accounting profit deducts from net operating revenues a figure for capital consumption based on historical cost and one or another conventional depreciation formula, whereas the economic estimate of capital consumption would be based upon the actual loss in value of the equipment to the enterprise. "Replacement cost" valuation of equipment is less incorrect than historical cost but still imperfect, since in many cases the economic value of equipment will have been degraded below replacement cost by the development of cheaper or more efficient machines. In any case the conventional depreciation formulas will be incorrect, since in economic principle what is desired is to recognize loss of value only as and if it occurs over time. As a secondary difference between economic and accounting profit statements, the former would exclude from profit a normal return on the owned equity in the business, while accounting profit deducts only the capital return paid out on borrowed funds.

use exceeds the aggregate cost at any output. If there is an excess, it will be greatest for the marginal-cost solution.

In either case, whether or not a profit is possible, a loss is incurred at the optimal output if a single price is charged. The many different economic methods developed to resolve the problem of full cost recovery for utilities exhibiting economies of scale are described in detail in Chapter Six.

b. Economies of Scope

The addition of the concept of “economies of scope” to that of economies of scale in defining natural monopoly was generated by the observation that most utilities were multi-product firms (Baumol, 1977). Economies of scope arise in situations where cost efficiencies are realized by producing a set of services due to cost complementarities. Cost complementarities refer to a situation where the production of one service means that another service can be provided at a lower cost than if produced separately. Typical water utilities offer the services of:

- potable water;
- water distribution;
- sewage treatment;
- sewers and storm sewers; and
- fire protection.

Water utilities also offer service to several distinct customer classes, many with distinct supply characteristics, predominantly in terms of the distribution and sewage grids and connections.

Given the observation of economies of scope, the definition of natural monopoly in economics has been altered to the concept of “sub-additivity”; simply expressed sub-additivity is a situation in which a single firm can provide a set of services at lower costs to all users compared to multiple firms providing separate services. Mathematically sub-additivity can be shown as;

$C(a+b) < C(a) + C(b)$, where $C(y)$ refers to the cost function and a and b refer to two distinct utility services.

Though seemingly simple, the concept of economies of scope has been profound in allowing for an assessment of utilities as the sum of their separate components. It has led, for example, to substantial restructuring in many utilities, notably in the separation of certain potentially competitive portions of utilities from those deemed basically uncompetitive. Most recently in Ontario, this has involved separating production from transmission in the electrical energy field. With respect to water utilities, the principal restructuring has occurred by separating distribution and sewage networks from treatment plants (both water and waste), with treatment plants more amenable to management under contract, as is common in Ontario with OCWA, or in the case of public/private franchise models of utility operation.

c. Natural Monopolies

Monopolies arise when a single firm or enterprise produces the entire output of a particular commodity. In the case of significant economies of scale and scope, conditions are such as to induce monopoly situations. In the municipal water utility industry, for example, capital requirements are so large that the co-existence of competing firms is infeasible, especially in the face of probable economic losses implied by socially optimal production. The conditions are quite common in industries that are both capital intensive and oriented to the provision of public goods and services.¹⁷

¹⁷ The issue of public goods is discussed in Section 2.2.3.

The existence of unregulated privately owned natural monopolies poses several problems for society. One problem involves probable economic efficiency losses. Stated simply, monopolies tend to maximize profits by limiting supply or raising prices (or both), because of the lack of competition. A second problem relates to the equity of income distribution under monopolistic pricing; basically, the existence of monopoly profits serves to transfer income from individuals being served by the utility to owners of the utility. The positive externalities related to public health inherent in water supply and effluent treatment comprises a third problem. Because an individual entrepreneur may not benefit proportionally to the social welfare generated by, for example, reduced morbidity or mortality, adequate water supply and treatment may generate no private interest in the absence of government regulation or provision. Even where private interest to build and manage water utilities is adequate, the utility may not supply water, or treat effluent according to adequate sanitation standards, or may limit or ration supply to some connections within a given geographic area in order to increase profitability.

These social considerations have given rise to three common forms of management for natural monopolies:

- government ownership;
- perpetual public utilities; and
- franchise public utilities.

Each type of ownership typically corresponds to a different type of governmental regulation. Government ownership corresponds to direct ownership of the utility by a particular level of government. This form of water utility management is the common form in Canada and Ontario, where ownership typically resides at the municipal level. A perpetual public utility exists where a private company holds a perpetual monopoly (or an indefinite franchise) in a given geographic area and submits proposals for rate structures and rate increases to elected or appointed regulatory authorities. This situation is reflected in current public water servicing in England. Franchise public utilities bid against other utilities to supply some or all services (i.e. production, distribution or both) to a community for a limited time contract, normally quite long in the case of water supply. This form of utility management is common in France. Chapter seven will address the economic aspects of utility organization and management.

2.2.3 Public Goods

Public goods can be defined in several ways. Standard textbook economics (see, for example, Tietenberg, 1996) defines a public good as one that both exhibits consumption indivisibilities, and is non-rival. A consumption indivisibility occurs when, once the resource is provided, even those who fail to pay for it cannot be excluded from the benefits it provides. Goods are non-rival when one person's consumption of the good in question does not diminish the amount available for others. Both of these characteristics tend to lead to public provision of the good, since private entrepreneurs are not able to capture sufficient revenues to cover their costs. While perfect examples of public goods are rare or non-existent, typical examples are street lighting or national defence. While in-stream water use also has substantial public goods characteristics, in the case of water utilities, potable water is both excludable¹⁸ and rival. Therefore, from the classic economic perspective, water utilities are not true public goods and, in practice, complete private ownership of water utilities is not uncommon. The United States and England provide examples of this private ownership situation.

A somewhat different perspective of public goods is provided by Crane (1993) who defines a public good as a good provided by government spending which is consumed collectively by members of the community. Integral to this definition is the concept of social perception of a "right" to a given service, where equity-based social decisions, rather than technical characteristics, determine the extent of

¹⁸ although exclusion may not be desirable on equity grounds.

excludability socially appropriate for certain goods (e.g. health care). In the case of potable running water one can observe that in Ontario, as in many other areas, the conscious exclusion of segments of the population from potable water is not politically or socially acceptable. Thus, water utilities do, in fact, have a strong public goods orientation, although, as noted above, they may not be perfect examples; similar services include roads, parks, public transit, education, and hospitals. This socially determined non-excludability generates a powerful incentive for intensive public involvement in the adequate provision of the good, which has in the past often assumed the form of public ownership. This has been the case with most public goods in Canada, including water utilities. However, a growing trend in terms of the provision of public goods is towards private sector involvement in their provision, on both ideological and efficiency grounds. Chapter seven highlights the central role played by cut-offs to connections for water supply by newly privatized water utilities in England in generating social unease with the privatization decision. This has led to a growing trend involving tighter regulation of those private utilities and their ability to disconnect consumers.

2.2.4 Externalities¹⁹

In many cases, private economic activities can have unintended consequences on people or firms not involved in these activities. One of the most common situations in which this occurs is when a party making decisions about resource use imposes damages on other parties. In other words, the decision-making party does not bear all of the consequences of its actions. Such consequences are referred to as "externalities," in recognition that they are external to the party (or parties) making the resource decision(s) in question.

Externality situations are very common. Suppose that a paper mill (with associated pulping operations) and a municipality use the same watercourse and are in close proximity to each other. Both use the watercourse for water supply, as well as for waste disposal purposes. Assume also that pollution control regulations are not in place. The paper mill uses large amounts of water and discharges large volumes of wastewater, in this case not being subject to pollution control regulations. The types, volumes, and strengths of this waste material are incompatible with the requirements for potable water of the municipality. Thus, extra costs are imposed upon the municipality as a result of the mill's operation. Because the mill does not bear the cost of extra water treatment downstream in the municipality, it is unlikely that its owners will be sensitive to those extra municipal costs in their financial planning and decision-making. As a result, it can be expected that the mill would discharge too much waste into the river, thereby preventing the achievement of optimal use of the resource.

In more general terms, externalities occur whenever a particular resource user is dependent upon his or her own actions plus the actions or activities of some other party. In the example, the municipality, in its water servicing decision-making, relies not only on itself, but also on decisions made by the paper mill operators, over which it has no control. The extra pollution in the river imposes an external cost on the municipality, which is unaccounted for by the paper mill operators.

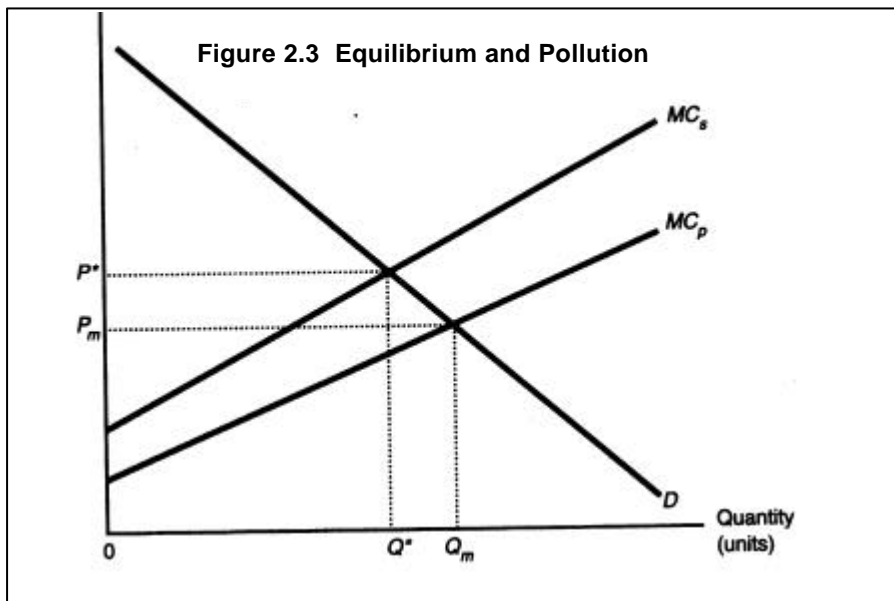
The effects of this externality can be shown by means of the following diagram (Figure 2.3), which depicts the market for paper mill products. As shown above, the production of paper involves the production of polluting materials as well as the mill's principal product. The demand for paper products is shown in the diagram by the demand curve D . The private marginal cost of producing paper (not including polluting control and damages) is represented by the marginal cost curve MC_p . Because society, in its definition of social costs, must take pollution control into account, the social marginal cost curve is represented by MC_s , which includes both the costs of paper production and the costs of downstream pollution to achieve potable water quality.

If the paper maker faces no public control on its emission of polluting materials, its optimal level of

¹⁹ This section draws on material in Tietenberg (1996).

production will be Q_m . In the private production setting, this is the level of production that maximizes the firm's profitability, or, in more technical terms optimizes its producer surplus²⁰. But clearly, from a social point of view, this is not the optimal production level, because from society's viewpoint, higher overall production costs are involved. These higher costs raise the (social) marginal cost curve to MC_s , and the optimal "social" equilibrium occurs at Q^* .

²⁰ The concept of producer surplus is outlined in Chapter 3.



With the help of this diagram, a number of observations are possible when externalities are not accounted for.

- Commodity output is larger than the socially optimal level.
- Excessive production of polluting material occurs.
- The prices for the goods produced by the polluting industry are too low.
- As long as the costs of pollution are external to the plant involved, no incentives exist to find alternatives to avoid the existing levels of pollution.
- There exist no incentives for pollution prevention, or for recycling or reuse of the polluting materials because discharging them into the receiving watercourse is exceptionally cheap (i.e., free).

The effects of these externalities are felt throughout the economy, because the market for the industrial good in question is larger than it would be under conditions in which the costs of pollution control are taken into account.

External effects may be either positive or negative. Situations where externalities are positive are commonly referred to as “external economies”; negative externalities as “external diseconomies”. Clearly, the situation described above would be considered an external diseconomy. However, external economies in general lead to less conflict. For example, the owner of a particularly attractive house enhances a neighbourhood and thereby provides benefits for his or her neighbours, as well as for passers-by (for which the latter two groups do not pay). A person maintaining a prairie pothole on his property provides nesting grounds for waterfowl, which, in turn, may benefit the hunting community. When these positive externalities exist, conditions opposite to those outlined above pertain. For example, production will tend to be too small, and prices for the good too high. Environmental externalities in the form of pollution, and its impact on water utilities, will be addressed in more detail in the next chapter.

2.2.5 A Renewable but Depletable Resource

In the natural resources field, it is traditional to view various resource materials as either renewable or non-renewable. The basic distinction between the two concepts involves the speed with which a given resource can be replenished. For non-renewable resources, the replenishment rate is very slow or zero. Crude petroleum is a good example. The replenishment rate of this resource can basically be viewed in terms of geological time – that is hundreds of millions of years. This means that current use of the resource depletes overall supplies, such that new reserves are essentially impossible to regenerate.²¹ On the other hand, renewable resources, as their name implies, are replenishable over relatively short periods of time. Water resources are typical. Over time, these resources are renewed frequently through the operation of the hydrologic cycle. Long-term perturbations to supplies may occur naturally or through human actions²², but supplies are essentially renewable over relatively short time periods (e.g., one year or less). Both renewable and non-renewable resources may be recyclable. The recovery of minerals (e.g., iron) from discarded products, and the recirculation of water by industry comprise examples of recyclability.

More recent thinking on the renewable-nonrenewable dichotomy has shown that a resource once thought of as renewable can be depleted to the point of becoming non-renewable. There are several instances the “renewable but depletable” nature of water resources in Ontario. These frequently, but not always, involve groundwater. The degradation of groundwater supplies in the Elmira and Manotick areas are examples of situations in which industrial contaminants have destroyed community water supplies. The accumulation of toxic substances in the bottom sediments of the Lower Great Lakes (Muir and Sudar, 1986) presents an example of partial depletion of the quality of even huge water bodies. The importance of recognizing the possibility of depleting renewable resources is that it is quite possible to ruin an essential local resource through neglect, inadequate regulation, and lack of enforcement of any regulations that might be in place.

2.2.6 Property Rights

As used in economics, a property right refers to a set of entitlements that define an owner's rights, privileges, and limitations for the use of a resource. Tietenberg (1996) and Bromley (1991) define a set of four basic characteristics promoting efficiency. These are²³:

- Universality – a condition in which all resources are privately owned, and all rights are fully specified
- Exclusivity – all benefits and costs resulting from resource ownership and use accrue solely to the owner, either directly or through sale to others
- Transferability – rights to resources are completely transferable between resource owners in voluntary exchange
- Enforceability – Rights are protected from involuntary seizure or encroachment by others

When these four conditions are established, owners have a strong incentive to use resources efficiently, because personal loss will ensue if they are not. While water resources generally fail to meet these conditions, in Canada, there are a limited number of small streams over which private property rights, as

²¹ This cannot be interpreted in terms of long-term scarcity in overall energy supplies. Scarcity is an entirely different concept that is not examined systematically in this report.

²² An example of human impacts on water availability might be occurrence of climate change due to accelerated rates of deposition of CO₂ into the atmosphere.

²³ Based on Tietenberg, 1996, p. 41.

defined above, have been established. In these cases, the owners develop the streams for high quality fishing experiences, and sell access rights to individuals for recreation. The streams are pollution free, and the ecosystem is improved so as to maximize the fishing experience.

In terms of the municipal water industry, the property rights issue is difficult to assess. Based on Demsetz (1967) the development of property rights stems from scarcity of the resource, where, as resources become scarce and valuable, competition for these resources generates interference in each individual's use of the resource, a good example of negative externalities. At some point, it becomes more efficient for society to regulate, modify, and refine the existing property rights system than to perpetuate them.

Based on the Canadian Constitution Act of 1982, the provinces own all in-stream water flowing within their borders. The federal government has jurisdiction in specific water-related areas, such as fisheries, navigation and international waters. For the most part, however, provinces have very important responsibilities for water resources.

As shown in the next chapter, water in Ontario is relatively abundant, and thus the property rights system is relatively simple. Ontario allocates in-stream water supply through a permitting system based on the Ontario Water Resources Act, with water permits currently managed by the Ministry of Natural Resources. A small, flat, but economically insignificant administrative fee is charged for processing a permit. In the context of Hayek's concept of prices as signals about resource values and scarcity, Ontario's permitting system implies that water is free, and of little value. This has profound, if unrecognized consequences for water management, as will emerge at several points throughout this report.

The various economic concepts are interrelated. To take but one example, the significant economies of scale implied by many publicly-oriented undertakings means that (a) economically efficient pricing may be difficult to establish, and (b) that natural monopolies tend to be common. When public goods, such as water services, are involved, the unregulated private provision of these goods may not fully serve the public interest, thereby requiring various forms of public intervention. The presence of negative externalities heightens the latter need. All of these conditions are present in the municipal water utility industry, posing significant challenges for economic analysis. This report centres on addressing these challenges.

CHAPTER 3: MUNICIPAL UTILITIES IN A WATER MANAGEMENT CONTEXT

3.1 Introduction

This chapter examines some broad economic issues in the water management field, from the perspective of in-situ, or instream, water management and its impact on water utilities. These issues are of central importance from the perspective of water utilities, particularly in terms of the demand for raw water and for the control and management of water pollution.

Water utilities are both affected by, and affect, the quality and quantity of in-stream water, because they both withdraw in-stream water as a basic input in the production of potable water, and discharge effluent to receiving water bodies. The costs to water utilities, and society, of the system of in-stream water management derive from both the costs of accessing and treating raw water, and costs imposed on utilities to treat their wastes adequately. This chapter will provide some documentation on the cost to water utilities associated with water quality on the intake side, while chapter six provides some estimates related to future upgrading costs related to effluent quality.

Water management issues have been traditionally dealt with using a quantity/quality dichotomy, which will be employed in this chapter. Having made this distinction, it should be emphasized that quantity and quality must also be dealt with in an integrated manner, as they are in much of the rest of the paper. In other words, we are using the quantity/quality dichotomy only as a means for organizing this discussion.

This chapter begins by discussing some relevant economic theory, combined with a very brief overview of selected legal issues, pertaining to water resource management in Ontario. Following this, major characteristics of water quantity and water quality are described. Each of these two sections is dealt with in approximately parallel fashion, proceeding from a discussion of basic physical factors, to a brief outline of current management approaches used in Ontario, and then to an economic assessment of these approaches. The chapter closes with a summary and synthesis of the material presented.

3.2 Economic Theory and Water Management

3.2.1 Legislated Rights: The Traditional Approach to Managing Water

Three basic economic concepts are central to in-stream water management: public goods, property rights and externalities; each of these concepts was discussed in the previous chapter. The public goods nature of in stream water stems from both the difficulty in excluding non-withdrawal users from its use (e.g. swimmers, boaters) and the Canadian preference for public ownership of natural resources, which developed in the late 19th and 20th centuries. This historical trend can be clearly seen from east to west in Canada, where the Atlantic provinces and Quebec exhibit a somewhat higher degree of private ownership of natural resources (e.g. a few small streams, standing timber), while in Ontario and Western Canada natural resources are almost exclusively owned by their respective provincial governments.

Given both this public ownership of in-stream water and the presence of significant, often negative, externalities, the essential economic problem for the government is that of efficiently and equitably allocating rights for the use of water. According to Demsetz (1967), the complexity of the efficient property rights allocation system depends on the scarcity of the resource, and thus the extent of the externalities generated by their unfettered use. As demonstrated by Hardin (1968), in his famous article entitled "The Tragedy of the Commons", and by many current examples, such as the destruction of the Northern cod stock off Newfoundland, the consequences of governmental inability to allocate property rights adequately in the presence of externalities can be catastrophic.

Governments generally use three main methods to allocate property rights: legal instruments (e.g. command and control regulations, water permits), prices (e.g. effluent charges) and tradable permits. Of these, the first is the generally preferred approach in Ontario, Canada, and much of the rest of the world.

Placed in the terms of chapter two, the usual response by public agencies to allocating access to (publicly owned) environmental resources has centred on regulating the use of these resources, based upon the use of tort law. For example, legally prescribed limits have been placed on municipal and industrial effluents discharges; these take the form of regulations mandating the types, volumes, and concentrations of wastes that can be emitted. In Ontario, these regulations are set for water-based effluents in order to achieve pre-determined water quality objectives. With respect to water quantity, regulations may take the form of approvals to withdraw water, limitations on water use during droughts, the establishment of "orders of precedence" for water use, and other such limitations. In Ontario, the provincial Permits to Take Water Programs comprises an example

In the context of this report, the most important characteristic of regulatory approaches, as currently used, is that the legal proscriptions seldom incorporate any economic incentives for compliance. Even in a strictly regulatory context, often the wrong means of assuring compliance have been used. For example, Rankin (1991) demonstrated that regulatory compliance could be achieved in two ways: through administratively set penalties (Rankin used fines set by the Workman's Compensation Board and an example); or through penalties requiring lengthy (and expensive) court proceeding (such as those set under provincial water quality regulations. Rankin found that compliance was achieved much more readily under a largely administrative system of assuring compliance than under the more complex court-based system. In most provinces, and also federally, the latter is the primary means of trying to enforce regulations, and compliance rates are generally low.

The main alternative, the use of economic instruments to assure compliance with society's environmental objectives (as discussed below), tends to be uncommon currently. Politicians and public administrators have worried about the political ramifications of using economic instruments in societies accustomed to essentially free rights to water, or legally based allocation of property rights to public facilities and watercourses. This is reflected in the usual public agency response to water pollution by trying to regulate the generation of polluting material, as outlined above. Industrialists may support neither regulation nor economic approaches because both options have impacts on the "bottom line" of firms, but, if forced, appear to support the former rather than the latter, possibly because they perceive that monitoring and enforcement may be difficult. The public also tends to support regulation for three basic reasons: a common perception of "legal certainty" engendered by legally enforceable "rights" to water; a reluctance to view environmental resources as marketable commodities; and a resistance to new forms of taxation.

3.2.2 Economic Approaches to Compliance: An Alternative

Given the nature of economics, and its emphasis on maximizing efficiency, much of the economics literature on property rights allocation advocates market or market-mimicking solutions, such as prices (e.g. effluent charges), as compared to legal instruments. In the economics literature over the past 40 years, effluent discharge fees, and more generally the entire idea of using “economic instruments” for pollution control has been hotly debated. Professional economists have been concerned largely with the possibilities of achieving perfect efficiency in the setting of fee levels, normally referred to as achieving “first-best” solutions, an aim that is both fraught with difficulties, and another reason for the lack of practical applications of economics-based approaches.

Nevertheless, some of these difficulties have been overcome by further research, and the use of economic instruments is slowly becoming more common, albeit almost invariably (and possibly advantageously) combined with regulatory approaches. This “hybrid” approach is motivated by the aim to bring some of the benefits of economic incentives and disincentives into play in allocating the resource. As shown in Chapter Two, and as borne out by experiences in many resource areas, the economic approach produces both long run social savings and technology change.

a. “First-Best” or Economically Efficient Approaches

The initial theoretical work dealing with externalities was undertaken by the welfare economist Pigou, who suggested that externalities could be dealt with by levying a charge on production high enough to bring about an efficient solution, with the price set at the level of the social damages or costs imposed by the externality. For example, in Figure 2.3 this amount would be $P^* - P_m$. This was the origin of the concept of pollution taxes or effluent discharge fees, but also applies to prices for water withdrawals. Given the imposition of prices by the owner of the common resource (e.g. the province), externalities would be internalized by firms and individuals, who would be provided incentives to reduce their externality-generating behaviour (e.g., excessive water demands, generation of large volumes of polluting materials, etc) in order to minimize their overall costs.

The major problem with Pigouvian prices involves determining the appropriate and acceptable value of damage associated with the externality in order to set the charge. This problem has been attenuated somewhat by quite recent research in resource economics, which has expanded the concept of value to include both market and non-market goods. This work includes non-market value concepts such as contingent valuation, which seeks to use concepts like bequest value (the value we place on leaving a clean environment to our children) and option value (the value we place on possibly visiting a pristine environment in the future) in placing hitherto “intangible” values into actual monetary terms. An example of non-market good valuation has recently involved the estimation of the costs associated with the recent, serious drinking water problem in Walkerton, as discussed later in this chapter. Despite improvements in non-market valuation, the determination of efficient Pigouvian prices remains a major problem in achieving “first-best” solutions.

b. “Second-Best” Approaches: Combining Efficiency and Practicality

To address the issue of achieving efficiency in pollution control, Baumol and Oates (1971, 1988) developed a pricing approach that we will refer to as “second-best.” In this context, the term “second best” means adopting solutions that sacrifice certain theoretical niceties offered by pure efficiency in favour of being able to use a particular instrument practically to meet a given end, in this case to achieve acceptable levels of pollution control. Second-best solutions normally reflect conditions in the “real world.” Baumol and Oates (1971) concluded, “... The adoption of a system of unit taxes ... has rarely proven feasible because of our inability to measure marginal social damages.” In place of “optimality”, and finding that the concept of charging systems had considerable merit, they explored the idea of an

iterative approach to setting unit effluent charge levels. This concept is a good example of combining two of the principles for dealing with the management issues posed by public goods, as outlined in chapter two – economic efficiency and practicality.

The Baumol-Oates formulation is an example of the “hybrid” of the two predominant environmental control approaches referred to above, in that it assumes that the public agency in charge first prescribes a set of standards, either based on best available technology, or water quality objectives, as is standard practice. Economic instruments in the form of effluent discharge fees are then used to facilitate compliance. Basically, the idea would be to set a “trial” charge level and, if it were found to be ineffective (e.g., in not procuring compliance with pre-determined regulations, or in proving too onerous), the charge would be adjusted appropriately. They stated that (1988):

“...The public authority can impose a system of charges that would, in effect, constitute a set of prices for the private use of social resources ... The charge (or prices) would be selected so as to achieve specific acceptability standards rather than attempting to base them on the unknown value of marginal net damages. For example, one might tax all installations emitting wastes ... at a rate of t (b) cents per gallon, where the tax rate t , paid by a particular polluter, would, for example, depend on b , the BOD value of the effluent, according to some fixed schedule. Each polluter would then be given a financial incentive to reduce the amount of effluent he discharges and to improve the quality of the discharge (that is, to reduce its BOD value. By setting the tax sufficiently high, the community would presumably be able to achieve whatever of purification of the river it desired. It might even be able to eliminate at least some types of industrial pollution altogether.

“ In marked contrast to an attempt at optimization, should iterative adjustments in tax rates prove desirable in charges and standards based approach, the necessary information would be easy to obtain. They require no data on costs or damages – only figure on current pollution levels. If the initial taxes did not reduce the pollution of the river sufficiently to satisfy the preset acceptability standards, one would simply raise the tax rates. Experience would soon permit the authorities to estimate the tax levels appropriate for the achievement of a target reduction in pollution.

“...Of course, such an iterative process is not costless. It means that some polluting firms and municipalities will have to modify their operations as tax rate are readjusted. At the very least, they should be warned in advance of the likelihood of such changes, so that they can build flexibility into their plant design ...but at any rate it is clear that, through the adjustment of tax rates, the public authorities can usually realize whatever standards of environmental quality have been selected.”

c. The “Double Dividend” Concept

Terkla (1984) developed a somewhat different perspective on environmental charges, based on the “efficiency value” of externality based tax revenues. The efficiency value of these charges was defined as “the reduction in excess burden resulting from the substitution of these revenues for current or future resources distorting taxes, such as those raised from federal corporate or personal income taxes.” The author found that these efficiency gains for two types of air pollutants ranged between \$0.63 billion and \$3.05 billion per year (\$1982) when substituted for federal personal income taxes, and between \$1 billion and \$4.87 billion when substituted for corporate income taxes. This led to a concept known as a “double dividend” resulting from the use of environmental charges. Whereas, prior to Terkla’s study, charges were recognized as being an economically efficient means of attaining pollution control, to these efficiencies had to be added additional sums to account for savings gained through substitution of effluent discharge fee revenues for other revenue, thereby raising the means of public agencies. This finding has important implications for a common criticism concerning environmental charges, namely

that these types of taxes act as a “drag” on industrial productivity. To the contrary, Terkla showed that “taxing the bads and not the goods” could improve overall social efficiency.

Effluent charges and withdrawal fees are relatively common in practice, being used primarily to raise revenues for administration, but also to provide incentives for achieving compliance with standards and regulations. The most advanced use of effluent charges is in Europe, notably in France, Germany, and the Netherlands, though they are also used in many U.S. states, as well as the province of British Columbia. A more detailed example of the French effluent charge system is provided in chapter seven.

d. Market Creation: The Concept of Tradable Permits

Dales (1968) provided a different perspective on the issue of externalities control. Whereas much of the work in this field had been devoted to various aspects of Pigouvian charges, Dales suggested that actual markets be created for allocation the capacity of the resources. In terms of water pollution externalities, the argument proceeds from the assumption that assimilative capacity of a water course can be quantified reasonably accurately²⁴, Dales’ market creation idea began with the idea the public agency managing the resource would issue permits up to the limit of this capacity, appropriately denominated. To qualify for the right to deposit waste, all activities (municipalities, industries, agriculture, etc.) within a watershed would require a discharge permit. The permits, following an initial distribution²⁵, would be marketable and tradable. Thus, firms could divest themselves of excess permits should they have relatively low pollution abatement costs. Conversely, firms requiring additional permits would be required to purchase them in the market. Over time, Dales’ system would tend towards an efficient allocation of the available assimilative capacity of the basin, as well of pollution control costs. An additional advantage here is that the public manager could “ratchet downward” the total emission permits available should an increase in water quality be chosen. The permit “market” would then respond efficiently, and achieve a new equilibrium, presumably at a higher permit price. In the same manner, new users could be accommodated as they located into the basin.

The Dales system was put forth in an attempt to achieve economic efficiency in the allocation of a watershed’s assimilative capacity. Paradoxically, it has seen limited application in the water resource context, but has received much more attention as a means for achieving air quality objectives. For example, tradeable permits lie at the heart of the Kyoto Protocol for reducing the emission of greenhouse gases.

In terms of water-based effluent control, the best-known applications of this system has been in the area adjacent to the Dillon Reservoir in Colorado, the Fox river basin in Wisconsin, and the Tar-Pimlico basin in North Carolina. These systems have generally been developed as systems trading between point source polluters (e.g. municipal utilities) and non-point polluters (e.g. farmers), recognizing the differences in cost of installing tertiary treatment in wastewater plants as compared to improved agricultural management practices regarding traditional pollutants, such as phosphorus. In the case of water abstraction, relatively advanced areas of water trading are the Murray- Darling basin in southern Australia, and the Southwestern United States. In both cases, following, Demsetz, more sophisticated permit trading systems are the result of real scarcity, for example, drought.

e. A Digression on the Coase Theorem, and Subsidies

Possibly the most famous explicit treatment of property rights allocation related to externalities was in the paper “The Problem of Social Cost” by Coase (1960), which examined the problem of which party should pay Pigouvian taxes. Should the cause of the externality compensate those affected, or should

²⁴ This has been found to be the case in South Africa, in the Delaware River Basin, in the Rhine Valley, and in many other areas. See Kneese (1977).

²⁵ Dales made clear that the initial distribution of permits could be by auction, by equal distribution, or by other means chosen by the public agency; initial distribution was not critical for the outcome of the proposed scheme.

the affected parties pay, in effect “bribe”, the polluters to reduce or eliminate their polluting behaviour? The conclusion reached by Coase was a rather remarkable one. Assuming that negotiation costs were negligible, that perfect information was available, and that the affected parties could negotiate freely with one another, courts could allocate the entitlement to the use of the water (and therefore the returns from the Pigouvian tax) to either party and efficiency would be achieved. In effect, the only difference between an efficient solution when either charges or bribes were used lies in the resulting income distribution, and its consequences for the use of resources (Mahler, 1974). This is basically a distributional, or equity, issue, not an economic one. This conclusion has come to be known as the *Coase theorem*. This theorem implies the relative equivalence of various instruments for addressing externalities, for example subsidies (bribes), taxes, and permits in terms of efficiency, though not equity.

However, in the vast majority of cases, information and transaction costs, which were assumed to be zero by Coase, are substantial. For example, a single firm or a small group of firms may be largely responsible for polluting the waters adjacent to a municipality or for exceeding the capacity of municipal treatment plants. On the other hand, the level of water pollution affects all of the residents of the municipality, including those who live downstream. Marshalling a consensus among all of the latter parties about the effects, and more especially about the levels and contribution to Coase’s “bribes”, even if theoretically possible would entail substantial costs, which brings the implicit unimportance of government in the Coase theorem into question. Thus government, given imperfect information and transaction costs, has a role to play in acting as the arbitrator of property rights. However, government has a relatively free choice in terms of instruments in terms of efficiency, with the choice of instrument often indicating an explicit or implicit property rights allocations made on equity grounds.

3.2.3 Summary

This section has contrasted two approaches to controlling water resource externalities. The first, and currently most common, primarily uses the force of law to set forth regulations on the allocation of rights to various characteristics of water resources – water supply, the bearing of pollution loadings, and so forth. The second, and much less common, approach attempts to marshal economic instruments to provide incentives for socially desirable behaviour (and, conversely, disincentives for socially undesirable behaviour.) The use of such economic instruments has proven successful in many areas. For example, it forms the basic rationale for using water pricing approaches to deal with several economic issues of municipal utility operation, a subject that is explored in detail in chapters four and six. In the end, however, and depending upon the situation being addressed, we suggest that combining the two approaches will be the most effective solution to dealing with municipal water issues in the future. Such an approach was exemplified in the outline of the Baumol-Oates research outlined above.

The various approaches and concepts described in this section can be used to evaluate current water management practices in Ontario in an economic context. We do this in the following two sections, which examine water quantity and water quality management. As outlined at the outset of the chapter, these evaluations proceed from a selection of basic facts describing current water quantity and quality, to a brief description of the predominant management approaches, and the to an economic evaluation of these approaches.

3.3 Water Quantity Issues in Ontario: An Overview²⁶

3.3.1 Water Supply: Demand Balancing

a. Water Supply

²⁶ This section is based on Tate (2002)

The atlas-like publication, *Water Quantity Resources of Ontario* (Ontario, 1984) contains an overview of facts on the physical water resources of the province. It divides the province into the five large hydrologic regions listed in Table 3.1. This table also shows a few of the important factors that give rise to Ontario's abundant water resources.

Table 3.1 Selected Water Quantity Characteristics by Hydrologic Region					
Region	Area (km ² x10 ³)	Mean Annual Precipitation (Mm) ¹	Mean Annual Runoff (mm) ²	Mean Annual Evapo-transpiration (mm) ²	% of Mean Annual Runoff Volume ³
Hudson-James Bay	571	650	300	350	59
Nelson River	122	660	210	450	9
Lakes Superior and Huron	175	850	400	450	21
Lakes Erie and Ontario	56	820	300	520	6
Ottawa River	49	810	310	500	5
Source: Ontario (1984). Interpolated from <i>ibid</i> , p. 16. Interpolated from <i>ibid</i> , p. 22. <i>ibid</i> , p. 26.					

From this table, several factors emerge. Ontario is a huge province in terms of area, encompassing about 973,000 square kilometers (km²). It has abundant precipitation, which falls both as rain and snow. Roughly 40 percent of precipitation comprises runoff, that portion of precipitation that reaches rivers and lakes from both surface runoff and groundwater flows. The remaining 60 percent is accounted for by evapotranspiration, and is thereby lost for direct human use. Finally, the table shows that over 68 percent of the flow (i.e. that in the Hudson-James Bay and the Nelson River regions) is toward the north, away from the populated areas of the province.

Not included in this table is the large storage capacity of Ontario's lakes, where annual run-off comprises less than 1% of storage capacity. The latter fact is important in assessing water availability. To draw an economic analogy, storage would comprise the total wealth of an individual or corporation. Annual runoff would be the annual interest earned on this capital. In order to preserve wealth, a useful objective is to live off the interest. If this analogy is accepted, it is the annual runoff (i.e., the flow), not the total storage (i.e., the stock), which is most important in assessing water availability. This flow criterion is used throughout the remainder of this section, even while recognizing that stocks of the resource are huge.

Annual surface water runoff, defined above as that portion of precipitation that reaches rivers and lakes from both surface runoff and groundwater flows, is quite variable from year-to-year in Ontario. Table 3.2 provides data on this flow variability.

Table 3.2 Annual Runoff (m³ per second) for Major Ontario Hydrologic Regions			
Region ¹	Reliable ²	Mean	High ³
Hudson-James Bay ⁴	3730	6000	8260
Winnipeg ⁵	380	760	1140
Great Lakes ⁶	2400	3070	3730
Ottawa	1390	1990	2590
Total	7910	11810	15720

Notes:

Source: Pearse et al. (1984, p.28). Note the contrast between the flows given here and the measurement of runoff in Ontario (1984), as sourced in Table 1. This is due to the different measurement units used, and not to any significant variations in basic data. All figures are rounded to the nearest 10.

1. The region names from Table 1 have been retained, with the exception of the Great Lakes, which are treated as a unit in the source for this table

2. Flows equalled or exceeded 19 years out of 20

3. Flows equalled or exceeded 1 year out of 20

4. Called "Northern Ontario" in the source for this table.

5. Measured at the Lake Winnipeg outlet point; accordingly the figures are biased upward due to the inclusion of (a minor amount of) runoff rising in Manitoba.

6. Includes all four Great Lakes, and measured at Cornwall

In addition to surface water runoff, varying proportions of Ontario's water supply are derived from groundwater. With respect to the latter, precise data are significantly harder to obtain because flow patterns are not well defined in many areas, and because measurements are discrete, not continuous. Groundwater yields generally are conditioned by geology. Hence, for example, in Northern and Eastern Ontario, where Pre-Cambrian and Ordovician rock formations are at or near the surface, groundwater yields tend to be low. On the other hand, where thick overburdens of glacial materials (e.g., those found in lacustrine deposits, kames, eskers, moraines, etc.), yields are much higher. Groundwater is important in the water supply of many smaller communities and almost all of rural Ontario, but may also supply portions of water supplies in larger urban centres. The largest urban area in the province that relies principally on groundwater is the Regional Municipality of Kitchener-Waterloo. Here, approximately 80% of municipal water supplies are abstracted from groundwater, the remainder from surface waters. Groundwater supplied municipalities are those most likely to suffer quantity constraints as local aquifers are exhausted and communities grow, leading to the need for increased capital investments to tap Ontario's relatively abundant surface water supplies.

b. Water Demand

In 1999, GeoEconomics produced a study of current and emerging water demands in the Canadian portion of the Great Lakes basin (Tate and Harris, 1999) for the International Joint Commission. The base year for this study was 1996, the last year for which measured water demand data are available in a systematic manner. Table 3.3 indicates that municipal water use is a relatively small proportion of total water intake, though a higher proportion of consumption, with a slightly negative growth rate of demand from 1986 to 1996.

Sector	Intake					Discharge				
	Year		1996	Annual Growth		Year		1996	Annual Growth	
	1972	1986		72/96	86/96	1972	1986		72/96	86/96
Agriculture	240	320	370	1.8%	1.5%	190	260	320	2.2%	2.1%
Mineral Extraction	70	100	110	1.9%	1.0%	0	0	20	0.0%	n.a.
Manufacturing	3990	3730	3610	-0.4%	-0.3%	150	100	190	1.2%	6.6%
Thermal Power	5630	19970	24000	6.2%	1.9%	100	200	210	3.1%	0.5%
Municipal	1310	1510	1360	0.2%	-1.0%	260	310	270	0.2%	-1.4%
Total	11240	25630	29450	4.1%	1.4%	700	870	1010	1.5%	1.5%

Based on a current project for the Ontario Ministry of Natural Resources GeoEconomics has developed the following table for the source of water use for public water supply. Groundwater constitutes a considerable proportion of total water use in certain river basins, notably the Huron and Erie basins.

c. Overall Water Balances

These numbers indicate that total water use represents less than one percent of Ontario's average annual runoff. Even for the Great Lakes, where most of Ontario's consumptive use takes place, the range of water consumption is still just over 1% of water availability. This has led many authors, including the present one, to call the Great Lakes a "water rich" environment. This finding must, of course, be qualified in light of basic geographical factors. For example, a community adjacent to a large river or lake will have a relatively smaller problem in acquiring water supplies than one located inland and away from large water bodies, or depending on groundwater supplies.

Sub-basin	Intake	(a) 1991			Groundwater Intake
		Consumption Low	Consumption Medium	Consumption High	
Lake Superior	76	4	8	11	4
Lake Huron	377	19	38	57	157
Lake Erie	836	42	84	125	231
Lake Ontario	2610	130	261	391	62
Great Lakes	3,899	195	390	585	453
(b) 1996					
Lake Superior	81	4	8	12	4
Lake Huron	393	20	39	59	171
Lake Erie	825	41	82	124	276
Lake Ontario	3,080	154	308	462	80
Great Lakes	4379	219	438	657	531

Source: Geoeconomics, 2002 (work in progress).

3.3.2 Managing Water Quantity in Ontario: Current Practices

Based on the Canadian Constitution Act of 1982, the provinces own all in-stream water flowing within their borders. Though the federal government has some jurisdiction in specific water-related areas, such as fisheries, navigation and international waters, the authority for water abstractions, such as municipal water intake, lies entirely with the province.

As shown above, water in Ontario in purely quantitative terms is relatively abundant. Correspondingly, the property rights allocation system employed is a simple one. Ontario allocates in-stream water supply through a permitting system based on the Ontario Water Resources Act. Under this Permits to Take Water program, currently managed by the Ministry of the Environment, all users withdrawing over 50 m³ per day from surface or groundwater sources require a water-taking permit. The permits are not transferable, and are issued for long-term periods. A small flat administrative fee is charged for processing a permit. In the context of Hayek's concept of prices as signals about resource values and scarcity, Ontario's permitting system implies that raw water has no marginal cost, implying that it has no economic value, possibly reflecting the relative lack of scarcity in Ontario related to water quantity, but also reflecting basic administrative and public perceptions outlined in Section 3.2.

3.3.4 A Brief Economic Assessment of Municipal Water Quantity Management

On an aggregate basis, Ontario has an abundant water supply, as demonstrated in Section 3.2. In situations of abundance, economic principles tend to be downplayed, because situations of scarcity are perceived as minor, or non-existent. In Ontario, for instance, the Permits to Take Water Program is the sole economics-related instrument in use, and, as shown above, levies only small fees on relatively large water users. These fees are irrelevant in terms of providing incentives for considering alternatives to excessive water use. In other words, current provincial practices governing the rights to use water are unrelated to economic concepts. For example, the practice of levying quasi-market-related royalties, such as are used in other areas of provincial natural resource management, is not employed with respect to water resources, although this concept has been examined by Tate and Rivers (1990) and Dupont and Renzetti (1999). Also, as indicated in Table 1.1, and as will be developed in more detail in Chapter Four, water prices in the province are very cheap, relative to the prices for other goods and services, as they are throughout Canada. Water prices in many OECD nations are substantially higher than they are in Canada, and Ontario (OECD, 1999).

Therefore, we find that Ontario, as a whole, does not employ economic concepts or practices in managing municipal water quantity issues. As will be indicated in Chapter Six, some municipalities employ water rate structures somewhat conducive to encouraging water (and therefore capital) conservation, but these are still in the minority (see Table 6.2 and the accompanying discussion).

3.4 Water Quality Issues in Ontario: An Overview

Providing a complete overview of water quality is considerably more complex than dealing with water quantity issues. Fundamentally, water quantity has only one parameter, volume (in either stock or flow terms), to be dealt with. On the other hand, water quality has many parameters, all of which can have

wide temporal or spatial variations. In preparing this section, we have selected only a handful of facts, which seem relevant to dealing with water quality in an economic and municipal utility context. The selection of facts is by no means comprehensive.

3.4.1 Effluent Discharge in Ontario

The material outlined here reviews the state of effluent treatment in Ontario, based on the Delcan (2001) for the Walkerton Inquiry. While water quality in the province is not without problems, important improvements in water quality have been made since the 1970s, notably decreases in Great Lakes nutrients (e.g. phosphorus), traditional pollutants (e.g. BOD, TSS), and certain toxic substances (e.g. DDT, PCBs).

a. The State of Waste Treatment

Six hundred waste treatment plants, both municipal and industrial, discharged treated wastewater into Ontario watercourses in 1998. Of these, 243 municipal systems were operated by OCWA, 207 by individual municipalities, and 163 by individual industries. Of the latter, 152 discharged treated process wastewater.

The total capacity of treatment plants in the province was 6.8 million cubic meters (MCM) per day. Typically, daily flows were substantially less than design capacity, because most plants have extra capacity built into them to allow for peak flows. The distribution of flow across the 450 municipal plants was heavily skewed toward the larger plants. For example, the 45 largest plants (i.e. 10% of the total) accounted for 81% of total capacity, and served 69% of the total population afforded waste treatment services. The average treatment capacity of these plants was 15,000 m³ per day, but the largest plant in the province, Ashbridges Bay in the Toronto area, had a capacity of just under 820,000 m³ per day. The remaining 90% of treatment plants in the province had capacities under 15,000 m³ per day. The impact of a treatment plant on ambient water quality is not necessarily related to its size, but rather to the size of the receiving water body. Thus, small plants discharging into small streams may have an overall local impact greater than a large plant discharging into a large water body.

Almost 4 MCM per day was discharged from industrial operations into Ontario surface waters. It is noteworthy that only the largest industrial operations had their own treatment plants. The vast majority of industrial plants in the province used municipal facilities; accordingly, some proportion the municipal treatment plant capacity cited above (6.8 MCM per day) was used for industrial wastes. (A simple estimate of this proportion would be 20%, the proportion of total municipal water demand accounted for by industry – see Table 4.1 below.) In the same manner as outlined for the municipal plants, industrial discharge was heavily skewed toward the large plants. The 45 largest facilities accounted for over 90% of the discharges from the 163 separate industrial waste treatment plants. The largest industrial discharger emitted 919,000 m³ per day, exceeding the total capacity of the largest municipal plant.

b. Municipal Discharges to Surface Waters

The Ontario Ministry of the Environment (MOE) collected data on 447 municipal waste treatment plants across the province. All but 11 of these discharged to surface waters, most of which are used also for potable water provision. Using the criteria of receiving the discharge from at least 5 facilities, or one discharge exceeding 20,000 m³ per day, 35 lakes or rivers in the province acted as repositories for municipal wastes. Based on treatment capacity, there were 156 “major” municipal dischargers to surface waters. These accounted for just under 90% of total discharge capacity of the municipal plants (i.e., 6.1 MCM per day out of 6.8 MCM per day. These major dischargers represented about 35% of the total number of municipal plants. These facts again demonstrate the skewed nature of municipal water discharges towards the larger plants.

Plants with relatively small capacities comprised under 10% of discharges to receiving waters, accounting for just under 220,000 m³ per day. As noted above, despite their relatively small size, these discharges may be locally important for water quality, because they discharge into smaller water bodies than most of the larger plants.

c. Municipal Treatment Technologies

Classifying waste treatment plants can be quite complex. Traditionally, a three-fold classification has been used. Primary treatment removes only wastes that can be filtered or settled without chemical action. Secondary treatment applies biological processes, such as activated sludge processes, to waste flows, typically following primary treatment. Tertiary treatment can take several forms to further process waste following secondary treatment. This is aimed at “effluent polishing” to achieve up to 99% of waste removal, an example being processes to remove phosphates and other soluble materials that may be too small to be removed by primary or secondary processes.

Many combinations of processes can be used – for example treating wastes from primary plants with tertiary processes. These, and other complexities, make it difficult to generalize as to the percentage of raw wastes being removed in Ontario municipal treatment plants. The three most widely used treatment processes in Ontario municipalities are extended aeration (used on 24% of total discharge), conventional (seasonal) lagoons (23%), and conventional activated sludge (21%). The widespread use of lagoons suggests the occurrence springtime “surges” in waste flows to receiving waters. Primary treatment was used to treat 5% of total waste volumes. A final point of importance here is the possibility that some industrial contaminants discharged to municipal treatment works may be incapable of treatment, and may enter receiving waters essentially untreated. The extent of this potential problem appears to be unknown, although the province has attempted to regulate it, with variable success rates in the past.

d. Municipal Waste Treatment Performance

Assessing the overall effectiveness of municipal waste treatment is very difficult to do comprehensively. Formally, the Province’s current approach is almost exclusively regulatory in nature and depends on two primary instruments: Provincial Water Quality Objectives (PWQO) and Certificates of Approval (CAs). The PWQO specify the nature of the materials and concentrations permitted in receiving waters. These determinations must be made on a case-by-case basis, requiring a determination of the assimilative capacity of a particular water body. This means, for example, that a plant discharging a relatively small quantity of waste into a relatively large water body may be allowed waste concentrations higher than a comparable discharger into a relatively small water body. The actual waste characteristics permitted to specific plants is embodied into CAs issued by MOE. The constituents of most concern are BOD, suspended solids, total phosphorous, nitrogen in various forms, and bacterial counts. Monitoring relies on samples taken by facility operators in the municipalities, which are then assessed in municipal laboratories (for the larger municipalities) or outside labs (for the smaller ones). The MOE local offices then determine whether plants are in or out of compliance with their CAs.

Table 3.5 shows the degree of compliance for the 447 plants overseen by MOE. This table is interesting for two major reasons: it indicates that records for almost 60% of the plants were inadequate to assess compliance with the CAs; and also that only 28% of plants met provincial standards. These data make it impossible here to assess the overall condition of municipal waste treatment across the province, but suggest that there may be many problems.

Status	# of Facilities	% of Total
In Compliance	127	28.4
Out of Compliance	28	6.3
No Discharge in 1998	9	2.0
No Surface Discharge	18	4.0
Insufficient Data	265	59.3
Total	447	100.0

Source: Delcan, 2001.

e. Storm Sewer Issues

In many of the older areas of urban Ontario, storm and sanitary sewers are combined, and treatment plants can experience storm surges or heavy springtime runoff due to snowmelt. These surges tax and may even overwhelm the flow capacities of treatment plants, causing the discharge of untreated sewage to receiving waters. At least, such occurrences degrade environmental quality; at worst, they may be harmful to public health.

This poses real challenges, both environmentally and financially. There is no question that storm water conveyance systems are required in urban areas, and thinking on this issue has gone through several stages. Before the effects of municipal sewage on ambient water quality were recognized, sanitary and storm sewers were combined, and all municipal sewage discharged directly to receiving water bodies. As the effort to improve water quality evolved, treatment plants were built to treat discharges prior to deposition into watercourses.

It was recognized early on that storm surges could be severe enough to require bypasses during high flow events, and thus to the escape of untreated sanitary waste. In 1998, 83 municipalities were recorded by MOE as having such bypasses, with 43 of these reported as having used these facilities. Seventeen of these reported primary bypasses (i.e., of all treatment facilities), with a total discharge of 1.9 MCM over 987 hours. Eighteen reported secondary bypasses (i.e., sewerage afforded primary treatment, but with secondary facilities bypassed), with a total discharge of 8.1 MCM over 1,645 hours. Eight plants recorded both primary (2.0 MCM [1383 hours]) and secondary (1.2 MCM [1,474 hours]). These amounts are about 0.5% of total annual provincial treatment capacity (13.2 MCM out of a total annual capacity of 2482 MCM. However, as the Delcan report points out, if the total volume of bypassed flows is converted to flow rates, the total bypass flow would be the 12th largest treatment plant in Ontario. This is probably a better context within which to place the storm water runoff issue.

A more recent development in the “debate” over storm water management called for the separation of storm and sanitary sewer, so that, during storm or high runoff events, at least sanitary waste would all be treated. Separate storm sewers would be used to convey the periodic storm water flows directly to receiving water bodies.

Most recently, it has been recognized that even storm waters may contain relatively large volumes of polluting materials. This philosophy calls for combined sanitary and storm sewers, but with treatment plant capacities high enough to treat combined flows. We cannot assess the relative costs of these various options in this report, for this task would require detailed engineering studies. However, this is an issue that will have to be faced in future financial planning.

f. Industrial Water Quality Issues

As noted above, the vast majority of industrial establishments discharge their wastes to municipal sewers. For the most part, these wastes can be treated in the municipal facilities. Incompatible wastes are most often pre-treated at the plants-of-origin, but either through accident or design, sometimes

wastes incompatible with conventional treatment plants may enter public sewer and treatment systems, and thence to receiving waters. The magnitude of this problem appears to be unknown for Ontario. However, it is beyond doubt that industrial operations increase the costs of municipal waste treatment.

Under the Municipal and Industrial Strategy for Abatement (MISA) program, MOE identified 163 large industrial establishments as major wastewater dischargers. These plants had their own intake and treatment facilities, received potable water from host municipalities, discharged sanitary wastes to municipal sewers, but treated and discharged their own wastewaters. Delcan estimated that the total discharge by these 163 plants totalled 9.7 MCM/day, but this is considered a substantial underestimate because once-through cooling water discharges from 13 electric power plants is omitted from the MISA data. A better estimate in industrial water use can be derived from Table 3.2 above. With this in mind, it is clear that industrial discharges greatly exceed municipal ones.

In general, it is process water (i.e., water that comes into contact with industrial semi-finished or finished products) that is of most concern for water quality,²⁷ and for this water, the MISA record is relatively complete. The total estimated process effluent by direct dischargers was just under 4 m³/day, or about 45% of what MISA considered the total volume of direct discharge in 1998. The iron and steel industry (7 of the 163 plants) accounted for about 38% of this volume, pulp and paper (25 plants) for 29% of the process effluent volume, and mining (43 plants) for 12 % of this waste.

As for the municipalities, MOE uses a regulatory approach to controlling industrial wastes, specifying limits on both volumes and concentrations of discharges. In contrast with the municipalities, the types of waste specified in the regulations varies substantially across industrial groups, and, in addition to BOD, suspended solids and nitrogen, includes a wide range of materials such as dioxin, many heavy metals, toluene, chloroform, and many others. Generally, bacterial contaminants are substantially less of a problem in industrial effluents, but this is offset by the need often to measure acute lethality using various aquatic organisms. Overall performance of industrial water treatment is measured against standards set by MOE, but is substantially more difficult to summarize because separate files must be maintained on each of the 163 establishments. For this reason, it is not possible here to assess overall industrial waste treatment performance. This would be a very difficult, time-consuming task.

g. Agricultural Runoff

Livestock manures and pesticides deposited onto agricultural lands are a growing source of concern in Ontario. For instance, the recent water contamination problem in Walkerton appears to have been caused by untreated cattle manure entering the adjacent public water source. Direct run-off and tile drains can convey contaminants directly into surface waters, while seepage into groundwater may contaminate both water supplies and the base-flow that eventually forms part of stream flow or other surface water bodies.

The Delcan report estimated that about 15.3 million dry kilograms per day of livestock waste was produced in Ontario in 1999. The BOD equivalent was about 5.4 million kilograms. This represented 4.5 times as much BOD as produced by the total population served by municipal waste treatment. About 85% of the latter is removed through waste treatment, while livestock runoff is essentially untreated. On the other hand, agricultural runoff is not discharged directly to receiving water, so the actual runoff is also a small (but unknown) proportion of total production.

²⁷ This is not to say that cooling water is not of concern, because directly discharged cooling waters can have substantial local environmental impacts, for example of aquatic biota. However, process water carries the bulk of environmentally harmful materials, and is of substantial greater concern from the waste treatment viewpoint.

Traditionally, animal waste disposal has been through application directly onto farmlands, where this nutrient-rich waste has been beneficial as fertilizer and soil conditioner. Application guidelines by the Ontario Ministry of Agriculture, Food, and Rural Affairs (OMAFRA), aimed partially at nutrient losses, from either animal waste or chemical fertilizers are designed to minimize eutrophication of surface waters. In contrast to chemical fertilizers, however, manures may also contain pathogenic organisms such as E. Coli O157:H7, which is highly pathogenic, and protozoa such as cryptosporidium, in addition to BOD.

Although not legally required in Ontario, animal wastes may be treated to lower BOD and pathogens prior to land application. Various treatment processes can be used. Compaction is used to reduce water content (and therefore weight and volume) prior to storage or transportation. This process will not reduce BOD or pathogens. Aerobic composting mixes the waste material with dry matter rich in carbon to produce a solid, which is placed in windrows, and turned over periodically, during which oxygen is transferred to organic-degrading micro-organisms. Properly operated, this process can reduce both the volume of manure and its pathogenic content. Disadvantages are losses of nitrogen through volatilization of ammonia, and the associated odour. Other treatment processes include both aerobic and anaerobic digestion/ All of these treatments are used in Ontario to varying degrees, but none are ideal in the sense of retaining nitrogen cost effectively, removing large proportion of BOD, and inactivating pathogens. No centralized records are kept on processes being used and in which locations.

3.4.2 The Economic Costs of Water Pollution in Ontario

Few systematic studies have been carried out in Ontario on the social costs of water pollution (i.e., pollution damages). Therefore, it is not possible to estimate a "global" number for these damages. However, it is possible to indicate (a) how such studies ought to be carried out, and (b) some recent estimates of damage in specific situations. We begin this section with a discussion of the few econometric studies that have been carried out on this issue. These are oriented toward estimating cost elasticities resulting from water pollution. The Renzetti study outlined below in detail is the only such study that has come to light for Ontario during our literature review, and its implicit cost functions may be appropriate for applied modeling for Ontario. Following this, an example of this costs incurred through groundwater contamination is outlined. This study is of the financial costs incurred to municipalities through the contamination of ground water by chlorinated solvents, a common form of toxic chemical. Finally, a study assesses the financial costs associated with the Walkerton tragedy.

a. Econometric Research Studies²⁸

Much of the econometric municipal water supply literature is directed at either obtaining econometric estimates of the degree of scale economies (Hayes, 1987; Kim, 1987; Boisvert and Schmit, 1997) or is concerned with testing for differences in efficiency between publicly and privately owned firms (Feigenbaum and Teeple, 1983; McGuire and Ohsfeldt, 1986; Teeple and Glycer, 1987a, 1987b; Raffiee et. al., 1993).

There are few econometric studies that explicitly consider the role of fluctuations in raw water quantity or quality on the cost of water supply. Feigenbaum and Teeple (1983) estimated a cost function for water delivery systems in which output was represented by a bundle of attributes. One attribute is an index of water treatment and this was found to have a positive and significant impact on costs. Teeple and Glycer (1987a, 1987b) also examined water supply costs using a sample of Californian water utilities

²⁸ These econometric studies are highly complex and mathematical. In particular, the Renzetti model contains very advanced analytical techniques. It is presented here in detail because it may be a source of econometric forecasting cost functions for SBC.

that use both self-supplied and purchased raw water. Self-supplied water and purchased water were found to be substitutes. Interestingly, however, the relationship between raw water supply and other inputs was found to depend on whether the raw water is self-supplied or purchased. For example, self-supplied raw water was found to be a substitute to capital but purchased raw water was a complement. Renzetti (1992b) included the level of water stored in mountain reservoirs when modeling the costs of supply for the Greater Vancouver Regional District wholesale supplier of water. Renzetti found that the marginal cost of supply is inversely related to water levels in mountain reservoirs.

More recently, Dearmont, McCarl and Tolman (1998) examined the impact of diminished water quality due to soil erosion on water treatment cost. The authors regressed average chemical treatment cost against total quantity of water treated, a measure of turbidity (adjusted for pH levels), a dummy that indicated whether water quality guidelines were exceeded and a measure of rainfall. The data were monthly observations over a three-year period for a small sample of Texas municipal water utilities. The estimated regression indicated that lowered water quality raised chemical treatment costs. In particular, the elasticities of chemical cost with respect to turbidity and rainfall were 0.27 and 1.74, respectively.

A recent econometric study in this area using Ontario data was conducted by Renzetti (2000). The study empirically modelled the technology of municipal water supply utilities and used this information to examine the impact of raw water availability and quality on supply costs. The estimation model was based on the assumption that municipal water utilities seek to minimize the cost of supplying a given quantity of output (Q). In their choices regarding inputs labour (L), energy (E) and capital (K) and raw water (W), utilities are constrained by prevailing input prices (p_L , p_E , p_K) and their technology as represented by the production function, $Q = f(L, E, K; W)$. The variable W represented the state of the raw water input as supplied by the environment and its construction is discussed below. Utilities were assumed to have no control over the value of W available to them and, as a result, W acts as a fixed input for the utility.

These assumptions imply that the utilities' technology may be represented by a restricted cost function (Halvorsen and Smith, 1986):

$$C = C(p_L, p_E, p_K, Q, W)$$

In the Renzetti study, cost was measured as each utility's annual expenditures on labour, energy and capital. Output (Q) was measured by the sum of annual recorded deliveries to residential, industrial, commercial and institutional customers. It was assumed that the state of the raw water, W, was a function of three distinct characteristics: mean water availability (W_{MN}), the variability of water availability (W_{SD}) and the level of contaminants present in the raw water (W_{QL}). These three variables were substituted in place of W in the cost function. The water utilities in the data set drew their raw water from either a lake or river. The W_{MN} variable was represented by the annual average of daily lake level or river flow rate observations, depending on whether the utility drew its raw water supply from a lake or river. The W_{SD} variable was measured by the standard deviation of daily lake level or river flow observations over the year 1991. The W_{QL} variable was an index composed of observations on six environmental contaminants: faecal coliform, lead, aluminium, benzene, PCB and trichlorobenzene.

There were 40 observations in the data set. Twenty-three of the utilities drew their raw water from a lake, and from a river. None of the utilities used groundwater as a significant source of supply. All of the utilities in the data set were self-supplied. That is, they did not purchase water from a regional wholesaler and, as a result, did not face an external price for raw water supplies.

A translog model having the following form was used to approximate the structure of the cost function:

$$\begin{aligned}
\ln C &= a_0 + \sum_i a_i \ln P_i + \sum_k c_k W_k + b_Q \ln Q \\
&+ \frac{1}{2} \left(\sum_i \sum_h a_{ih} \ln P_i \cdot \ln P_h + \sum_i \sum_k d_{ik} \ln P_i \cdot W_k + b_{QQ} (\ln Q)^2 \right. \\
&+ \left. \sum_k c_{kk} (W_k)^2 \right) + \sum_i d_{Qi} \ln Q_i \cdot \ln P_i \\
&+ c_D D + \sum_k c_{Qk} \ln Q \cdot W_k + e_t
\end{aligned}$$

for $i, h = K, L, E$ and $k = MN, SD, QL$

A dummy variable was included to differentiate between water utilities using a lake (D=1) or river (D=0) for its raw water supply. The cost function and the labour and capital share equations were estimated jointly using an iterative Zellner SUR procedure with linear homogeneity and symmetry imposed.

The findings relative to water quality impacts were all statistically significant. The estimated elasticity of cost with respect to the availability of raw water supplies (e_{MN}) indicated that decreases in the average availability of raw water raised costs by a relatively small margin (a 1% increase in average water availability led to only a 0.07% decrease in costs). This may occur because water utilities had already installed collection systems (such as deep intake pipes) that allowed them to operate independently of lake and river depths. It is likely that this elasticity would have been larger if utilities that rely on groundwater sources had been included in the sample.

In contrast, increases in the variability of raw water supplies raised costs and had a larger proportional impact than do changes in the average water availability. The estimated elasticity with respect to the variability of raw water supply (e_{SD}) indicated that a 1% change in the standard deviation of water supplies leads to almost a 1% increase in total cost. The difference between the estimated values of e_{MN} and e_{SD} may have occurred because increases in the variability in raw water supplies have more significant impacts on system design. For example, these may necessitate expanding the scale of a number of the components of the utility supply network such as reservoirs, pumps, storage tanks and treatment facilities.

Finally, the cost elasticity of raw water quality (e_{QL}) was positive. This estimate indicates that increases in the level of contaminants that make up the water quality index increase the costs of supplying a given quantity of output.

b. Economic Damages from Chlorinated Solvents

Chlorinated solvents are a commonly used group of toxic chemicals, predominately used the dry cleaning industry or as industrial solvents. Based on a survey of municipalities carried out for Environment Canada (1995) the total financial cost to water utilities from groundwater pollution by chlorinated solvents totalled almost \$200 million nationally, and just under \$125 million for Ontario.

These figures represent expenditures made from 1985 until 1995. In addition, annual O&M costs were estimated at \$2.9 million. At an annual discount rate of 5%, the discounted stream of costs forecast into the future (capital and operating) total about \$38 million. Table 3.6 provides the details of these pollution cost estimates.

Site	Total Costs (\$10 ³)	Annual O&M (\$10 ³)
Smithville Municipal Water Supply	29,800	300
Bracebridge (Leader Spring) Municipal Water Supply	20,700	150
Gloucester Landfill	8,200	1,000
Manotick	5,500	25
Killiloe	4,200	40
Bristol Aerospace Plant, Rockwood	1,100	25
Highway 2, Brockville	300	15
McDougll Landfill, Brockville	200	10
Chemical Road, Brockville (Private Water Supply)	110	15
Region of Waterloo	1,171	45
Region of York	144	30
Simcoe Municipal Water Supply	200	15
Ingersoll Municipal Water Supply	130	25
Fergus Municipal and Private Water Supply	239	6
Robert Street Well Field, Penetanguishene	940	20
Angus Private Water Supply	360	10
Total of Measured Sites (60% total estimated damages)	73,294	1,731
Provincial Total	122,156	2,885
Annual 20-year costs discounted at 5%	-----	38,000

Source: Environment Canada (1995)

c. The Costs of the Walkerton Drinking Water Contamination Problem

In May, 2000, the town of Walkerton, Ontario experienced a major drinking water contamination problem, in which 7 persons lost their lives, and some 2,300 persons became ill to varying degrees. This incident is dealt with in great detail in the recently released report of the Walkerton Inquiry. One of the studies commissioned for the Inquiry dealt with the economic costs of this incident (Livernois, 2002). This study estimated both the tangible and the intangible costs incurred resulting from this incident.

Tangible, or direct, costs are those that can be measured by survey, or other reporting means. The study used detailed survey and interview methods to measure these costs (Table 3.7). The study estimated total tangible costs at \$64.5 million.

Table 3.7 Summary of Estimated Economic Costs of Walkerton		
Cost Category	Cost Sub-category	Cost Estimate (\$000)
Households	Walkerton households	6,876
	Non Walkerton households	40
	Property values	1,106
	Drinking water	4,167
Business	Direct business costs	1,460
	Lost productivity	1,234
Individual health care	Hospital stays	438
	Opportunity cost of time spent in hospital	51
	Physician's visits	99
	Long-term health costs	2,498
Health related	Epidemiological studies	212
	Helicopter ambulance costs	160
Public health agencies	Local public health unit	2,775
	Assistance to regional health unit	375
	Chief coroner's office	509
	Walkerton health study	5,000
Facility-related costs	Water testing, laboratory, and auditing	645
	Remediation and repair	9,222
Other regional municipal costs		6,549
Legal and other	Walkerton Inquiry	9,000
	Private legal costs	1,000
	Other agency costs	11,110
TOTAL COSTS		64,257
Source: Livernois (2002, p. 3)		

The intangible costs are much more difficult and controversial to estimate, for they involve, to a large degree estimating the value of a human life and the intangible loss of welfare due to illness. Livernois himself acknowledge that it is pointless to try to estimate such a values. However, he correctly pointed out that it was quite different to ask what society would be willing to pay to reduce the risk of a lost life, and the incidence of a serious illness. Existing economic models set the range of the former at between \$5 million and \$15 million, and the latter at \$15,000 per illness. Livernois used an average value of \$8 million per statistical life and \$15,000 per illness in the study. This placed the total intangible value of this incident at \$91 million. This is an example of the use of contingent valuation methods referred to earlier in the chapter.

Combining the tangible and intangible costs, the value of preventing future water tragedies like Walkerton was estimated at \$155 million. Livernois stated that, "The report is intended to provide a context for future action, allowing people to weight the costs and benefits of investing resources into providing safer drinking water. To take these estimates one step further, on might observe, on average, that the benefits (i.e., the forgone damages) of providing safe drinking water would be \$31,000 per capita for each inhabitant of Walkerton."

3.4.3 Managing Water Quality in Ontario: Current Practices

The relatively larger extent of externalities related to water quality can be seen in the relatively greater complexity of the property allocation methods used to address water quality issues in Ontario. In terms of water quality both the provincial and federal governments have jurisdiction, with federal authority stemming primarily from constitutional powers over fisheries and international waters, such as the Great Lakes.

The basic provincial method for addressing water quality from point sources, such as industrial or municipal wastewater treatment plants is primarily regulatory in nature and depends on two primary instruments: the PWQOs and the CAs. The PWQOs specify the nature of the materials and concentrations permitted in receiving waters. These determinations must be made on a case-by-case basis, requiring a determination of the assimilative capacity of a particular water body. As noted earlier, a plant discharging a relatively small quantity of waste into a relatively large water body may be allowed waste concentrations higher than a comparable discharger into a relatively small water body. The actual waste characteristics permitted to specific plants are embodied into CAs issued by the MOE, and are dependent on the nature of the effluent and the quality of the receiving environment.

The federal government also regulates certain types of point source effluent discharge under the Fisheries Act, where the much quoted section 36 states that "... no person shall deposit or permit the deposit of a deleterious substance of any type in water frequented by fish..." The federal regulations are based on variations of best available technology, with regulatory standards based on either quantity of pollutants related to production, or concentrations of pollutants in effluent. The sectors regulated are the pulp and paper, chlor-alkali, refining, chemical and metal mining sectors.

The federal government also regulates toxic substances under the Canadian Environmental Protection Act (CEPA). CEPA requires the federal government to assess substances for toxicity to either human health or the environment. An initial priority substances list of 44 chemicals was assessed in 1994, with a second PSL list of 22 chemicals was as of 2001. Actions under CEPA to manage those substances deemed toxic can be slow to develop, for example for chlorinated municipal wastewater effluent, a substance declared toxic under the first PSL, management plans were still being assessed as of 2002.

Finally, the International Great Lakes Water quality agreements, jointly implemented by Ontario and the federal governments, have also addressed water quality related externalities, with a focus on nutrients, such as phosphates, and toxic substances. A later focus of the 1987 management strategy was on joint remedial action plans associated with regional "hot spots". A new Canada Ontario agreement (COA) is currently in draft stage, the last expiring in 2000.

3.4.4 A Brief Economic Assessment of Water Quality Management

Subsidies have played a substantial role in the COA, and other wastewater clean up programs, but have declined somewhat over time, and, as noted earlier, the bulk of waste treatment funding originating locally. Effluent charges or tradable permit systems have not been used in Ontario in managing water quality.

Non-point source pollution, such as agricultural pollution, is beginning to be addressed in Ontario. A small subsidy program, the voluntary Environmental Farm Plan, funded by the federal government under the COA provided subsidies (\$ 3.9 million) for the development of the plans in the agricultural sector up till 2000, while an earlier subsidy (\$ 60 million) from OMAFRA (Clean Up Rural Beaches) was terminated in 1995.

The availability of abundant water resources has a very significant water quality dimension. Abundant water quantity by no means implies good water quality. On the contrary, abundant water supplies in many areas has often meant that water bodies became a (free) sink for all types of waste materials from every part of the economy – industry, agriculture, municipalities, and so forth. The adequate treatment of waste is actually quite expensive.

The alternative to waste treatment for many centuries has been deposition of untreated wastes into publicly owned watercourses free of charge. In general, when any enterprise faces an economic decision, there is a strong tendency to choose the least expensive alternative – in this case, the deposition of untreated waste, and over time, this externalities problem resulted in the deterioration of the quality of water resources.

The approach to controlling water quality in Ontario is largely still a regulatory one. The concept of using economic instruments to support this task, while undoubtedly well known, is not employed provincially. As illustrated in section 3.2, the incentives offered by the use of various economic instruments have proven successful in securing desired behaviour with respect to pollution generation and release. However, in Ontario, this type of policy instrument remains to be developed for general use.

3.5 Summary

This chapter has examined some broad economic issues in the water management field, from the perspective of in-site, or in-stream, water use management and its impact on water utilities. Water utilities are both affected by, and affect, the quality and quantity of in-stream water, as they both withdraw in-stream water as a basic input in the production of potable water, and discharge effluent to receiving water bodies.

The concepts of public goods, property rights and externalities are central economic concepts in addressing water resource management. The principal economic problem is that of efficiently allocating the property rights to a publicly owned resource in order to address externalities associated with its use. Under the Canadian constitution provinces own the in-stream water, while the federal government has authority over fisheries, navigable water and international waters.

The basic forms of possible property rights presented are allocation through legal “rights”, such as regulation, charges and subsidies, or some form of tradable permit. Though economists have concentrated on the efficiency value of charges and permits, common practice favours the use of legal instruments, such as effluent regulation or water rights, and subsidies. Nevertheless, the use of effluent charges is common with examples being France, Germany, and the Netherlands. Tradable effluent permit systems have been piloted in the United States, while form of water rights trading are being piloted in Australia, and the South West United States.

In terms of water quantity, Ontario is relatively water rich, though local conditions associated with groundwater may generate occasional scarcity. Accordingly, the system for allocating property rights to water for intake is relatively simple, consisting of provincially administered water taking permits, and small administrative fee.

Water quality problems are more prevalent, and include point source and non-point source pollution. Some common sources of water pollution impacting on water utilities include traditional pollutants (turbidity, suspended matter), agricultural run-off, and toxic substances. The relative extent and complexity of water pollution externalities give rise to a more complex system of property rights allocation, including separate provincial and federal regulation of point source effluent, federal regulation of toxic substances, and joint provincial-federal cooperation through the international Great Lakes Water Quality Agreement. The use of substantial subsidies was also historically associated with these programs, with subsidies declining radically in the recent past. Non-point source pollution, such as agricultural run-off, is not specifically addressed in Ontario.

Municipal water utilities are regulated by the provincial permits system (Certificate of Approval) based on ambient water quality (Ontario water quality objectives) and best available technology. There were 600 waste treatment plants, both municipal and industrial, discharged treated wastewater into Ontario

watercourses in 1998. Of these, 243 municipal systems were operated by the Ontario Clean Water Agency, 207 by individual municipalities, and 163 by individual industries. Of the latter, 152 discharged treated process wastewater. Municipal effluent accounts for approximately 40 % of total discharge of total effluent volume regulated, compared to 60 % for industry. BOD production associated with agricultural operations (e.g. manure) represents 5.4 times that of municipal waste, is untreated and unregulated, but is not discharged directly into receiving bodies. A recent report by the province assessing compliance of municipal wastewater plants with their CAs indicates that 28 % of plants were in compliance, and that almost 60% of the plants had inadequate data to assess compliance with the CAs.

Theoretical econometric studies related to pollution indicate a relationship between raw water quality parameters, such as turbidity or variability of supply, and increased water supply treatment costs. Some cost estimates related to the costs of pollution to water utilities are presented, including economic costs of \$ 155 million from Walkerton, and \$ 125 million over ten years for groundwater contamination from chlorinated solvents. Possible upgrading costs related to water supply and effluent is found in chapter six.

The relatively newly developed field of water resource economics provides several insights into the water management problems currently being experienced. The use of economic instruments, such as effluent charges or tradeable effluent permits, appears to offer new alternatives, and, in particular, supports for future management. The main lesson from this chapter is that current legal, regulatory and engineering-based approaches can be more fully supported by economics-based approaches in order to achieve sustainable water management.

CHAPTER 4: THE DEMAND FOR MUNICIPAL WATER SERVICES

4.1 Introduction

The size of a municipal water utility system depends fundamentally on the demands²⁹ imposed upon it. Determining demand is a complex undertaking, often considered secondary to determining supply, such as new capital and operating requirements. New supply is often estimated based on simple demand forecasts, where the projected demand is viewed as requirements that must be met under all conditions. The alternative view, which is examined in this report, is that demands are variable. They depend on a set of conditions, such as price, income, and others, which may assume many different values, and which can actually be influenced substantially by policy choices.

4.1.1 Municipal Water Demand in an Environmental Context³⁰

The sources of a municipal water supply, as well as its demand, vary widely across time and space. The availability, quality, and cost of a water supply system depends upon the geographic juxtaposition of source, plants, and consumers. To secure and deliver an acceptable water supply, in terms of potability, adequacy and reliability, municipal water managers must overcome the friction of distance at some specified cost. Temporal variations in distribution reinforce spatial variations because both must be “evened out” to meet demand. For example, the highest demand on a water source may occur at its lowest time of water availability. These environmental and distributional factors impose significant costs on water systems and influence the capital intensity of the municipal water industry. The irony here is that many current financial and economic practices actually exacerbate these natural frictions.

The demand for municipal water servicing and its fluctuations over time and space have been the subject of much discussion, both in the formal, published literature, as well as in water-related public agencies. Taken as a whole, and as reflected in this report, these studies reflect the state of the art in explaining the levels and variations in municipal water demand. The problem of explaining municipal water demands is also frequently hampered by the absence of reliable, systematic data, and this has led to a fairly large variety of general approaches and, at times, contradictory results.

It is generally agreed the municipal water use is a response to a number of conditions (e.g., physical conditions, economic factors, managerial decisions, and others), all of which vary over time and space to influence the patterns and levels of water demand. In a manner typical of all scientific research, these conditions have been reduced to a number of factors, the effects of which may be isolated, while holding as many other variables constant. This scientific approach has several advantages, according to Grima:

- It is suited to quantitative analysis so that under certain conditions conclusions correct within stated confidence limits may be stated.
- It explores some of the interrelationships between municipal water use and selected determinants (some of which are physical, others social and economic; some are subject to management influence, others are not)

²⁹ It should be noted here that the terms “use” and “demand” have different technical meanings. “Use” refers to the general activity of using water for whatever purpose. “Demand” is used to denote the economic relationship between price and quantity used. In common parlance, the two terms are often used interchangeably. In this report, we have tried to use the term “demand” only in the economic context

³⁰ Some of the discussion in this section has been adapted from Grima’s (1972) work.

- Statements of wide applicability may be made about the effects of variables that are not unique to points in time and space
- The tentative conclusions not only explore the nature of human behaviour in using a basic natural resource, but also have implications for improving the quality of life by better planning

This chapter explores the nature of municipal water demands in detail. It begins with a brief overview of water demand patterns in Ontario municipalities between 1989 and 1999, in order to indicate the magnitude, and principal sources of this demand. It then examines a number of factors underlying municipal water demands. These discussions focus on factors that help determine the volume and timing of demands by the various major water users within a municipality. This discussion is followed by an overview of the academic research in the water demand field, with an emphasis on residential water demands, which have received considerably more attention than commercial, publicly supplied industrial, or institutional demands. The emphasis here is on the comparability of the reported results, in terms of content and methodology, and the validity of the hypothesized relationships. This overview will demonstrate the ways in which the variables described in the immediately preceding section have been used in attempting to explain the factors underlying some of the major municipal water demands. The chapter closes with a brief discussion of water demand management. This discussion is a logical means by which to integrate the concepts of this chapter, using them to suggest an approach that alters the traditional focus of water management slightly, from an almost sole emphasis on supplying services, to one that gives equal emphasis to demand factors in attempting to achieve greater levels of efficiency in water servicing expenditures.

4.2 Municipal Water Demands in Ontario

4.2.1 Total Water Use

Municipal water demand in the Ontario municipalities surveyed rose from 141 billion cubic meters per month (BCM) in 1989 to 160 BCM in 1999 (Table 4.1). This table breaks total demand into three user categories: residential, commercial, and industrial plus other. The residential users class dominated the municipal water use for Ontario municipalities for the entire decade, accounting for approximately 50 to 60 percent of total usage. The table also demonstrates that the residential category of usage was more dominant in the smaller municipalities than in the larger ones. This is to be expected, because larger municipalities have larger commercial and industrial economic bases than small ones. This is reflected in the table by the drop in the percentage of water use accounted for by the residential category from between (approximately) 70 and 80 percent in the smaller communities (i.e., the first to size category) to between (approximately) 50 and 60 percent in the larger ones. In terms of total demand, commercial and industrial water uses accounted for almost the same percentage of municipal water use throughout the decade of study, around 20 percent. It is noteworthy that industrial establishments often negotiate “bulk water contracts” with municipalities, which supply water at much cheaper rates than for other users. Often, cheap water is used as a factor in attracting new industries to an area. This practice, known as “promotional” water pricing is a factor that will elevate water usage³¹. The extent of this practice is unknown at this point, but it is an important factor that should be kept in mind when discussing municipal water rates.

³¹ Contracts are normally confidential, and therefore difficult to obtain, but the practice is well-known among municipal water managers. Also, when decreasing block rate pricing structures (see Section 4.5.5 for a description of rate structuring) are used, industrial water demand often falls into the upper blocks (i.e., the lowest priced blocks) of rate schedules. This is another good indicator of promotional water pricing.

Table 4.1 Water Use in Ontario Municipalities, by User Group and Population Size Group							
('000 cubic meters per month)							
Size Group ('000 persons)	Volume of Usage				Percentage of Usage		
	Residential	Commercial	Industrial + Other	Total	Residential	Commercial	Industrial + Other
1989							
1 - 5	4,345	605	582	5532	79%	11%	11%
5 - 10	4,521	627	917	6065	75%	10%	15%
10 - 50	18,036	3,468	4,336	25840	70%	13%	17%
50 - 100	11,664	3,491	3,642	18797	62%	19%	19%
>100	43,857	20,403	21,154	85413	51%	24%	25%
Total	82,422	28,595	30,630	141647	58%	20%	22%
1991							
1 - 5	3,655	814	533	5002	73%	16%	11%
5 - 10	4,043	894	948	5885	69%	15%	16%
10 - 50	15,358	2,891	4,115	22364	69%	13%	18%
50 - 100	22,245	7,782	6,211	36239	61%	21%	17%
>100	26,298	9,657	9,098	45053	58%	21%	20%
Total	71,600	22,038	20,905	114543	63%	19%	18%
1994							
1 - 5	3,415	828	672	4914	69%	17%	14%
5 - 10	3,685	903	1,523	6111	60%	15%	25%
10 - 50	11,473	3,215	7,274	21961	52%	15%	33%
50 - 100	9,434	3,756	5,145	18335	51%	20%	28%
>100	38,607	21,204	25,028	84838	46%	25%	30%
Total	66,614	29,906	39,641	136161	49%	22%	29%
1996							
1 - 5	3,630	766	424	4819	75%	16%	9%
5 - 10	3,718	875	789	5382	69%	16%	15%
10 - 50	14,239	3,640	3,994	21873	65%	17%	18%
50 - 100	10,840	4,271	2,713	17824	61%	24%	15%
>100	55,302	25,654	18,746	99702	55%	26%	19%
Total	87,728	35,206	26,667	149601	59%	24%	18%
1999							
1 - 5	2,165	539	204	2908	74%	19%	7%
5 - 10	2,940	635	728	4303	68%	15%	17%
10 - 50	15,809	3,918	3,688	23416	68%	17%	16%
50 - 100	10,269	3,227	2,786	16282	63%	20%	17%
>100	65,361	26,790	20,858	113009	58%	24%	18%
Total	96,544	35,109	28,266	159919	60%	22%	18%

Table 4.2 places total municipal water demand into the context of Ontario-wide water use, showing that the municipalities accounted for only 5.7% of the total in 1999. The thermal power sector (conventional + nuclear) accounted for the bulk of Ontario's water demand (82% in 1996)³².

Table 4.2 Ontario Municipal Water Intake in the Context of Total Provincial Water Use, 1999 ('000 M³ per month and [10⁶m³ per year]) ¹			
Area	1996	1999 ⁶	2001 ⁵
Great Lakes Basin ²	2,450 [29,430]	2,570 [30,850]	2,650 [31,800]
Ontario ³	2,690 [32,260]	2,820 [30,810]	2,900 [34,840]
Ontario Municipal ^{4,7}	150 [1,800]	160 [1,920]	170 [1,980]

Notes:
¹Figures are rounded to the nearest 10. Annual amounts are shown because they correspond to previously published data.
²Figures for 1996 and 2001 derived from Tate and Harris, 1999, pp A1.6 and A1.7
³Ontario's total water use for 1996 was 9.6% greater than that for the Great Lakes. See Table 2.5.
⁴The volume for 1999 is taken from Table 4.1. For 1996 and 2001, it was assumed that the proportion of municipal: total intake was the same as for 1999.
⁵This volume represents the trend line projection from Tate and Harris, 1996, p. A1.7
⁶Total water use for 1999 was interpolated from the 1996 and 2001 volumes.
⁷For 1996 and 2001, the municipal water use is smaller than the volumes shown here, because, in the source report, the municipal component of other water uses (mainly the manufacturing sector) was included in the respective sectors. For 1999, the municipal component was separated from the other sectors, and consolidated into a more comprehensive municipal sector. The choice of methodology depends on the nature of the study being carried out.

4.3 Nature and Characteristics of Municipal Water Demand

4.3.1 Present Context

Municipal water demand can be examined from several different perspectives. For example, many water demand analyses have been conducted within single urban areas (e.g., Grima, 1972). In these studies, data are collected for individual dwelling units, and regression models built to explain the variations in water demand among those units. In these studies, variables have to be measured, of course, at the individual dwelling unit level. Another class of study looks at variation across wider geographic areas, and for these, measurements would be made at the municipality level. Variables might be quite different in the latter class of studies. For example, precipitation variations would be more important for these types of investigations, whereas they would not be a factor within individual municipalities. Other studies focus on seasonal variation, and a still different set of variable would be used. This is not to say that variables differ completely between the different types of studies. For example, water pricing practices or levels appear to be constant among the various types of research.

For descriptive purposes in this report, the main interest is at the cross-municipal level. In the following section, the conceptual model has been specified to apply to this level, and variables specified accordingly.

³² 1996 was the last year for which measured data were available on a multi-sectoral basis.

4.3.2 A Conceptual Model

Levels of demand for municipal water supplies and waste treatment are a joint function of several factors. These can be shown in conceptual form as follows:³³

$$Q_i, Q_w = f (P, PO, I, M, RS, PK, US, HHC, PPt, NR, C)$$

Where Q_i = the demand for water intake, or water supply, including peaking characteristics;

Q_w = the demand for wastewater treatment.

P = the price of water per unit (e.g., per m³).

PO = prices of other commodities

I = Average income levels in the community.

RS = type of rate structure in use

M = the degree of metering of water services.

US = urban structure

HHC = household characteristics

PPt = Average annual precipitation

NR = % of non-revenue water (e.g., fire flows, leakage, etc.)

C = presence/absence of an effective water conservation program.

4.4 Dependent Variables: Water Supply and Waste Treatment Demands

The quantity of water intake is the most common meaning of the term “water demand,” and is the principal determinant of the size of water systems. As noted in the specification of this variable, it refers to both the gross quantity of intake and to its peaking characteristics. The quantity of wastewater generated is conventionally assumed to be a function of water supplied into the system. The waste treatment variable is seldom measured in Canadian municipalities, except possibly at the intake points of treatment plants, and is considered essentially to be equal to intake minus a certain percentage of use that is consumed, for example through leakage, evaporation, and so forth. The consumptive use is higher in the summer than in the other months of the year. Accordingly, to be more accurate, the above function could be divided into summer and non-summer components. However, because we are not conducting an econometric exercise, the function can remain as stated. Finally, the waste treatment dependent variable as stated here does not include any qualitative parameters, although this could be incorporated readily.

4.4.1 Basic Requirements: Water Supply

³³ We have shown these in a format similar to that used in multiple regression analysis for convenience only, and following a format used by Bower (1966) for industrial water demand. The conceptual mode should probably not be used in the form stated for actual research studies. As pointed out by one of our collaborators, this conceptualization departs from the more formal econometric models that have been employed in the academic literature, and may include characteristics that might more appropriately be considered as a sub-set of another variable. Also, this conceptual model is set in the context of residential water demand, a commercial or industrial demand model would appear somewhat differently. Finally, the model as stated is static one, but could easily be made dynamic by specifying a time variable.

Intake quantity has several basic requirements. It must be available, at potable quality and in sufficient volumes, 24 hours per day, every day of the year. Many provinces in Canada, including Ontario, have legally mandated requirements of minimum acceptable drinking water quality. Also, there is usually close cooperation between local water supply authorities and public health officials. Pressures within water systems must be suitable to serve all customers, regardless of their location in the municipality, and including those located in the tallest buildings. Requirements also vary between water sectors within municipalities, particularly with regard to peaking patterns. Several general points concerning peaking characteristics are discussed below, as well as in chapter six on water rate setting. The main point here is that each of these physical requirements imposes additional capital and operating costs on water utilities.

a. Peaking

Within any municipality, there is a range of water demand types – from residential to industrial – and each has its own pattern of demand for both water and sewer services. Contrasting residential and industrial user groups will illustrate this point.

Residential demand is influenced by many end uses, which divide generally into two classes – indoor (washing, cooking, bathing, etc.), and outdoor (lawn and garden irrigation, vehicle washing, etc.). Each of these uses has its own unique time pattern of use, but two principal characteristics emerge: first a daily peaking of use, normally centred on evening meal times, and second, a seasonal peaking due to higher demands during the summer. Both of these features are important considerations in setting water rates, which will be dealt with in chapter six

Industrial demand has a different set of characteristics. Industrial use, based on a reasonably sized, modern, urban area can be extremely complex. However, a few characteristics are common. First, many large industries provide their own water supplies, because it is often cheaper to do so than to pay municipal rates. The general principle is that the larger the industry, the more likely it is to obtain the bulk of its water from self-supplied sources. Second, even large industries may rely on municipal water utilities for drinking and sanitary purposes. Third, industries that require large amounts of potable water (e.g., food and beverage producers) tend to obtain relative higher proportions of their water supplies from public sources, again to minimize costs (Tate and Scharf, 1995). Fourth, as already noted, many industries within a municipality negotiate “bulk water contracts” with the municipality, thereby avoiding the normal water rates. This reflects a frequent practice by municipalities to use “promotional” or low water rates as one means of maintaining or augmenting its industrial base. Finally, the daily and seasonal water use patterns by industry tend not to display the substantial peaking characteristics of the residential sector (Tate and Scharf, 1995).

Peaking characteristics are of key importance in determining the capacity of water servicing systems, because system capacity is partly oriented toward meeting peak, not average, demands. An important question here relates to the ways in which peak demands can be reduced, because if substantial reductions can be achieved, corresponding decreases in design capacities and financial requirements can be achieved.

The issue of peak load pricing and its importance for rate design is addressed in chapter six. This discussion will show that charging higher water prices during the summer, peak period months than during the winter, or off-peak months, offers a means of lowering peak period water demands, thereby lowering overall costs in the long run. To anticipate this discussion somewhat, seasonal rates are a potentially effective means for realizing more efficient use of scarce water resources when demands on a water utility's system vary systematically across seasons. Their primary advantage is that they provide to consumers an accurate signal of the cost of consumption, including the cost of capacity during peak use periods. In this regard, seasonal rates have several advantages over more traditional approaches to pricing capacity, including:

- Peak period users are made responsible for capacity costs pay for those costs. Traditional approaches typically spread these costs over all periods, which increases inefficiency and decreases equity. It increases inefficiency by underpricing water service in the peak period, thus encouraging too much consumption, and by overpricing water service in the off-peak period, thus encouraging too little consumption. It decreases equity because off-peak users, by paying a share of cost for which that they are not causally responsible, thereby implicitly subsidize the consumption of peak users.
- All uses during the peak period are recognized as contributing to and are charged for the cost of meeting the peak.
- Ideally, seasonal rates will reflect the full cost of capacity required to meet the peak rather than just that portion in excess of average demand. Traditional approaches, on the other hand, go to great lengths to identify whether capacity is meeting average demand or peak demand requirements. In fact, current capacity in any specific situation jointly meets both types of use, and causal responsibility for costs depends on the relative magnitudes of peak and off-peak demands. In cases where the differential between the two demands is large, the peak period will bear responsibility for the costs.

4.4.2 Wastewater Treatment

The contrast between residential and industrial water users is just as important on the waste treatment side of the water use cycle. Municipal waste treatment systems are set up primarily to treat domestic biological wastes, and demand patterns are roughly similar to those outlined above for water supply. The main difference is that there tends to be a substantially smaller seasonal peaking characteristic for residential waste, because much of the increased summer demand is for lawn and garden irrigation, which does not enter the wastewater collection and treatment system. With respect to industry, many of the same patterns as outlined above hold true for industrial waste treatment demand. The most noteworthy difference lies in the fact that industries may emit wastes significantly in excess of domestic wastes, in terms of both volume and strength. Also, some industries may emit toxic wastes that cannot be treated in normal municipal treatment systems. All of the factors must enter the water rate making calculus if water pricing is to be effective. This significantly increases the complexity of rate making.

Waste treatment requirements are also important in system design. In Ontario, the Great Lakes Water Quality Agreement, and several IJC-mandated Remedial Action Plans have required most municipalities in the basin to install at least secondary, and usually tertiary waste treatment systems. The provincial Ministry of the Environment also oversees the requirements for wastewater treatment. As in the case of water supply, waste treatment is very capital intensive, and forms a major component of local public works expenditures.

Waste treatment demands are usually oriented toward treating organic wastes generated by the municipal population. However, as the size of a municipality increases, industrial and commercial activity also rises. While the wastes from commercial activities are normally consistent with those of the population in general, the same cannot be said of many industries. The latter may cause two serious types of externalities – excessive discharge of both organic materials and materials that cannot normally be treated in municipal treatment plants – examples of the latter being heavy metals, phenolic compounds, etc. In the former case, additional capacity will be required to handle the increased organic waste loadings; in this case, extra-strength sewer surcharges may be an effective management instrument. In the latter case, there will be a requirement for pre-treatment facilities at the source plants to remove incompatible waste materials, and for monitoring to ensure compliance. In the case of large industries, treatment of many wastes will be handled internally, with completely separate facilities. In

these cases, it may be possible that only sanitary wastes will be deposited into municipal collection and treatment systems.

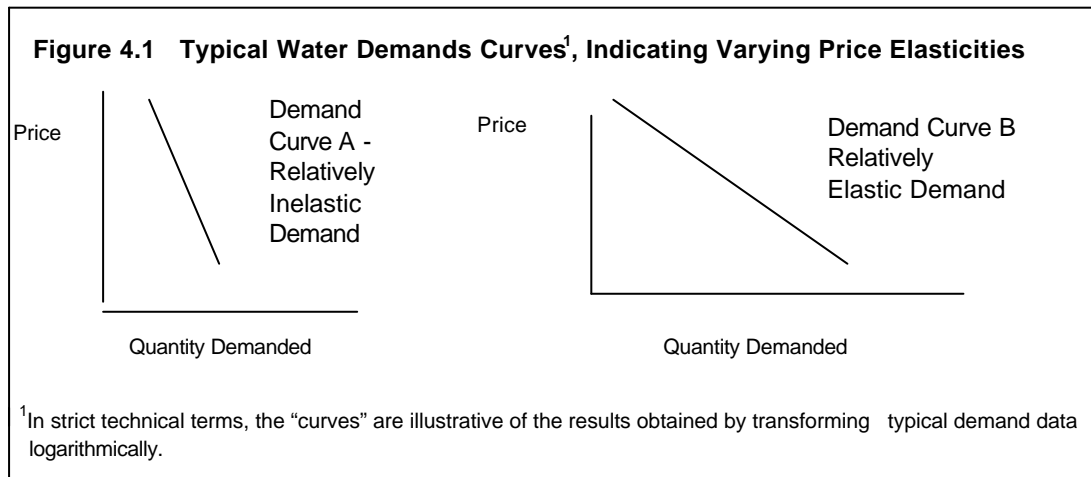
4.5 Independent Variables: Economic Factors

4.5.1 Water Prices

One theme of this report is that water demand is inversely related to price, in a manner similar to most other goods and services. The aim in this section is to treat this issue systematically.

Prices perform two essential roles in a market system: rationing and production motivation. Rationing is necessary since scarcity precludes both the satisfaction of all needs and the unlimited production of goods and services. Goods and services must be rationed to consumers and factors of production must be rationed to producers. The price system allows bidding for scarce goods and services and factors of production, thereby ensuring goods and services are allocated to the highest valued users and that factors of production are allocated to that use where they will bring the largest return. Prices perform their production-motivating roles by indicating what consumers are willing to pay. Producers react to this, directing their production to those products most profitable to them. In the terms discussed in chapter two, issues of pricing relate strongly to objectives of achieving economic efficiency.

The relationship between the price of water and the quantity demanded can be shown using the economic technique called the demand curve. Two such demand curves are shown in Figure 4.1. These typical water demand curves demonstrate that price and quantity are inversely related. The variations in the slope of the two demand curves are important in defining the way in which price and water demand are related, and will be discussed below.



There are, however, qualifications that must be placed on the use of demand curves in the water resource context. Demand curves were developed by economists in the context of “perfect markets”. In such markets a number of strict conditions apply. For example, no individual consumer or no individual producer is large enough to dominate the market. Also, it is assumed that each consumer and producer has perfect knowledge of both the price and the cost conditions pertained in the market. While these conditions approximate the conditions that pertain for most goods and services in a market economy, there are notable exceptions. For example, in a situation of monopoly, an individual producer is able to dominate the market. More related to water resources is the existence of externalities, which was

discussed in chapter two. Therefore, the case of water resources violates many of the conditions which define perfectly competitive markets, and accordingly the relationship may be somewhat weakened. Most important of all is that in most water markets there is an absence of competition among producers of the service, in this case municipal water and wastewater services. Despite these qualifications, this economic model still provides some useful insights for water demand management.

For resources with public goods characteristics, the absence of competition has two effects on prices. First, in most cases, pricing has simply never been a management consideration, and resources have been free. Second, water prices are set by administrative fiat. For example, municipal councils, or comparable public bodies, set the levels of prices charged for publicly supplied water. Theoretically, it is possible that each consumer could be assessed a quasi-competitive price for consuming the resource in question, in this case water services. Hartwick and Oleweiler (1986, p. 392) term these “personalized prices”. They continue by stating that:

“... there is (a) no natural mechanism causing these correct prices emerge ... , and (b) if consumers got together to register their preferences, the transaction costs of getting together even if the correct preferences were to emerge, would be very high. Thus we see that:

- *No private market can supply a public good such as clean air (or water).³⁴*
- *Some form of government intervention is required.*

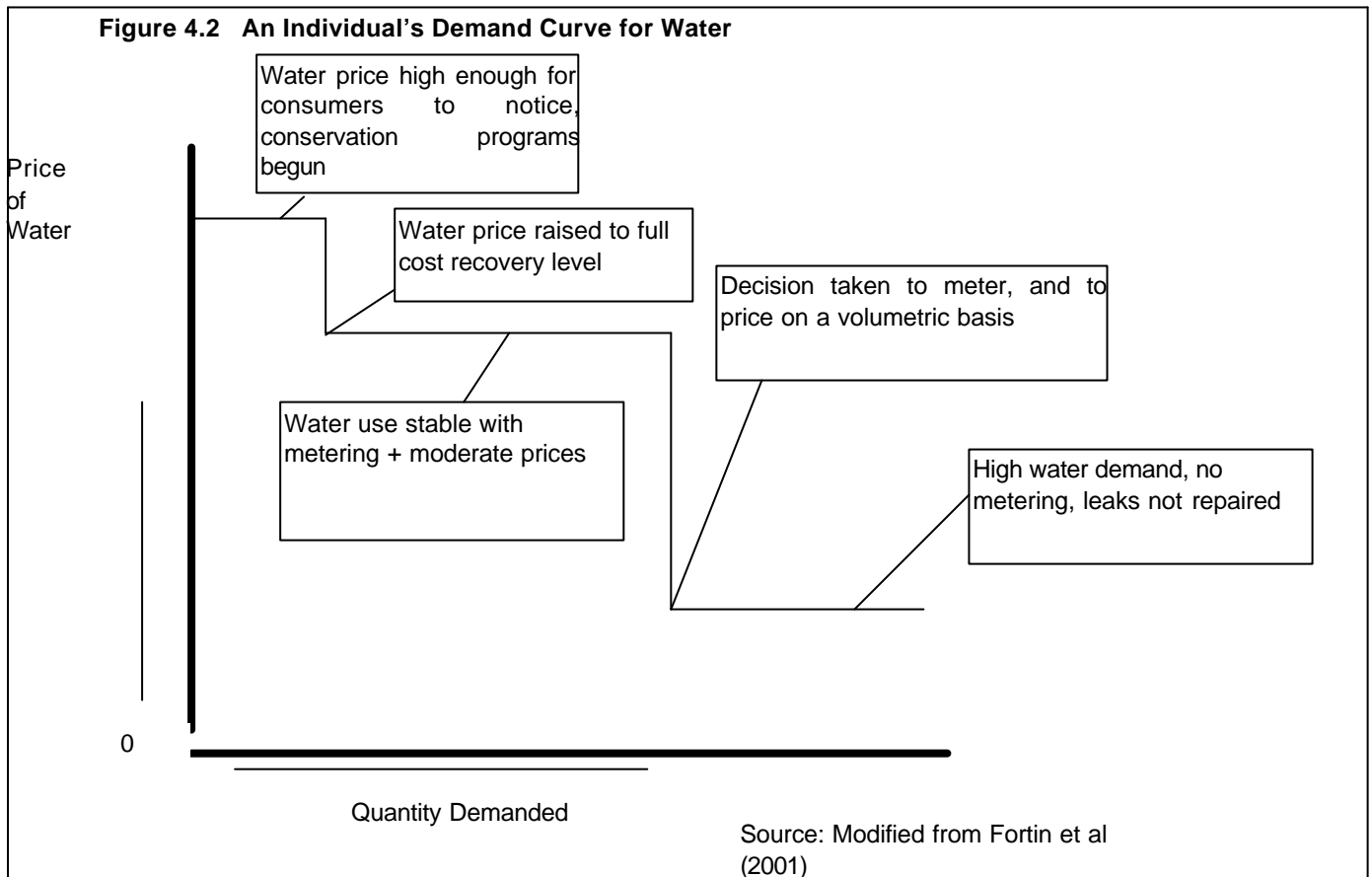
Even if the government does regulate the supply of clean air (or water) and charges a tax to individuals for that air, it is quite likely than the supply will be different from the optimum because individuals have an incentive to hide their true preferences.”

Thus, at a technical level, it is unlikely that administered prices will achieve the economic efficiency that is the outcome of a more competitive situation³⁵. Nevertheless, it is equally true that even administratively set prices can provide environmental resources users with varying degrees of incentives for resource conservation. In other words, the efficiency of environmental resource allocation will likely be improved under any type of pricing regime, even though complete efficiency is likely unattainable. This view reflects the second-best approach taken by Baumol and Oates (1988) in the context of achieving Pareto-optimality with respect to effluent discharge fees, as outlined in chapter three. It will be recalled that they suggested that fees did not necessarily have to be optimal to have an incentive effect with respect to reducing pollution. In the same manner, it is our contention that volumetrically-based water pricing does not have to be optimal to achieve a substantial lowering of demand. In other words, “second-best” solutions are both easier to develop and effective in promoting better financial management of water utilities.

³⁴ Bracketed portion added.

³⁵ We address the competitive issue in Chapter 7.

The recently published study by Fortin *et al.* (2001) contains a more accurate illustration of the nature of demand curves for water (Figure 4.2).³⁶ In contrast to the idealized demand curves for water shown



earlier, the demand curve shown here is a stepped function of water use against price, indicating that water demand responds to significant events, but is otherwise quite stable. Three “zones” of water demand are shown. Working from the right side of the figure, the first zone has low prices and relatively high demands. This would be the case in municipalities that are unmetered, with flat rate pricing in effect. The second zone has lower water demands, for example following the decision to meter water use, and to introduce volumetric pricing. As indicated in Figure 4.1 volumetric pricing will induce a lower water demand. Demand will remain approximately in this zone until prices increase sufficiently that consumers begin to notice them (e.g., the full cost recovery levels). This is indicated in the diagram with demand rising into the third zone. At this point, aggressive water conservation programs may be instituted to further curtail water demand. This figure comprises a demand curve for a single individual. Many individuals working within identical conditions will make decisions at different price levels (and quantities demanded), and thus the “market” demand curve will slowly change to one with a more “integrated” and “smooth” nature like those shown earlier.

³⁶ We modified the diagram somewhat to follow conventional economic practice, whereby demand curves trend downward to the right, as in Figure 4.1.

4.5.2 Other Input Prices

Economic theory predicts that water demand is a function of both its own price and all other prices faced by a household. In particular, we would expect the prices of sewage treatment, energy, and water related capital could each influence residential water demand. This issue is addressed in detail in Section 4.8

4.5.3 Income Levels

Research has found that water demand varies directly with income. This make sense because higher income level users will acquire more water using appliances and larger properties, on average than lower income ones. This variable is of less interest than water price because utility operators, for the most part, are unable to influence income levels, whereas they can have a direct impact on water price levels.

4.5.4 Price and Income Elasticities: A Digression

It was shown above (Figure 4.1) that demand curves can have different slopes. The economic concept of price **elasticity of demand** reflects this characteristic of demand curves. Price elasticity of demand is defined as the percentage change of quantity divided by the percentage change in price. Price elasticities are usually interpreted in the context of the percentage change in quantity demanded resulting from a given percentage change in price. Values of price elasticities³⁷ can range between 0 and infinity. In economics, elasticities with values between 0 and 1 are termed inelastic demands, those with elasticities exactly equal to 1 are called unitary demands, will those with values greater than 1 are termed elastic demands.

The elasticity of a demand curve actually reflects the availability of substitute for a good or service under consideration. The fewer substitutes that are available, the more inelastic the demand curve. Conversely, as the number of substitutes increases, the price elasticity of demand also increases. Also, price elasticity tends to rise as price rises, chiefly because increased prices make alternatives to current commodity use more competitive. Thus, one may infer that if and when water price rise, price elasticities of demand will also rise as the search for substitutes intensifies.

This discussion of price elasticity of demand has direct relevance to municipal water demand. A large number of research studies have shown that the demand for indoor water usage is relatively inelastic (see appendix to this chapter), and have relatively narrow ranges for give situations. The value of (absolute) water demand for indoor usage has been found to fall in the range of -0.3 to -0.4. In other words, a 10% increase in price will lower the demand for indoor use by 3% to 4 %. This shows that there are few substitutes for indoor water uses, for example, sanitary usage, drinking, cooking, etc. On the other hand, elasticities for outdoor water usage tends to be greater. Reflecting the fact that there are more substitutes available for outdoor water use, the values of outdoor elasticities tend to range between -0.5 and -0.7. In other words a 10% increase in price will lead to a 5% to 7% decrease in demand. Other factors influence price elasticity values. They tend to be lower when prices are low, because, fundamentally, there is no incentive to search for substitutes or to modify water-using practices. Also, price elasticities tend to be smaller in the short run than in the long run. In the latter, changes in capital

³⁷ Because price and demand are inversely related, elasticities usually assume negative mathematical values. The convention in economics is to denote elasticity values in absolute terms, and to refer to relatively elastic demand curves as having "greater values" than relatively inelastic ones, even though, in strict mathematical terms, this may be incorrect.

stock can occur, thus expanding the range of options available. On average, it can be stated that most municipal water demands are inelastic in nature, and behave in ways that are suggested by theory.

Industrial water demands tend to have elasticities somewhat higher than those for residential use because there are more alternatives available for industrial water usage (e.g., increases in recirculation, changes in production processes, etc.; see Bower, 1966). Interestingly, the demand for irrigation water in the agricultural sector is even more elastic than for either municipal or industrial demands.

In addition to the price elasticity of demand, demand can also vary according to income level. In general, the greater the income level of an individual, or in our case the average income of a community, the greater the demand for water. As income levels increase, water users tend to employ more water using equipment in their homes, businesses, etc. For example, dishwashers and swimming pools become more common. In contrast to price elasticity, the income: water demand relationship is a positive one. A range of actual income elasticities is also shown in the appendix to this chapter, where it will be noted that these measures are widely variant from one study to another.

Of the two elasticity concepts the price elasticity is the most useful because it allows planners and financial personnel to estimate both the water demand effects of price rises and the effects of such changes on revenue generation. In general, for inelastic demands (e.g., as in Curve A in Figure 4.1) increased revenues will be generated by price rises, because demand reductions are relatively small. For elastic demands, the opposite revenue impacts would occur. This price-revenue relationship is important for municipalities, because, with relatively inelastic demands, as is the case for most municipal uses, price increases will generally increase revenue.

In modeling municipal water demands, and more especially in designing municipal water rates, price elasticity of demand is an important consideration. This theme will be discussed again in chapter six. Selection of a demand elasticity for a particular situation depends on several factors:

- the range of customer types in the municipality. If this range is a narrow one, such as a largely residential community, elasticity values will largely reflect those in this one class; more complex urban structures require consideration of other user categories;
- the lower the price of water, the lower the elasticity values; thus, absolute price level is an important consideration;
- The mix between indoor and outdoor uses is important, because elasticity values are greater for the latter.

4.6 Independent Variables: Structural Factors

4.6.1 Water Metering

Water metering has a potentially large impact on water demand. For example, Kellow (1970) measured the impact of metering in two western Canadian cities, Calgary and Edmonton. At that time, Calgary was completely unmetered while Edmonton was completely metered. The per capita water use in Edmonton was about 50% of the level in Calgary.

This result needs some discussion. Metering by itself need have no impact on the demand for water since meters are merely a device attached to a water system to measure the intake of water for a set of customers. It has been argued (Grima, 1972) that metering by itself, unaccompanied by price changes will have no impact on water demand. On the other hand, some studies have shown that the psychological effect of knowing that the amount of water used is being measured may have up to a 20% effect in decreasing water demand. This psychological effect may be temporary, for it has been found

that where water metering is accompanied by price changes, water demand “rebounds” to near its original level. It is the capability to permit better pricing practices that is the real payoff from water metering.

A rigorous comparison of metered versus unmetered water demand was developed by the Johns Hopkins Residential Water Use Project in the mid 1960’s. A sample of 39 areas in the United States, each including a relatively homogenous residential district, was master metered during the period of 1961 to 1965. Demands were recorded every 15 minutes. The study included 10 metered areas with public sewers; 5 with septic tanks; and 5 apartment building areas without individual residential metering. The results of this study are shown in table 4.3.

	Gallons per day per dwelling unit	
	Metered	Flat rate
Annual average		
Leakage and waste	25	36
Indoor use	247	236
Sprinkling	186	420
Total household water use	458	692
Peak hour	2481	5170

Source: Linaweaver, Beebe, and Kriven (1966); as cited in Billings and Jones (1996).

Apart from showing a huge per capita water demand in all cases, but particularly in the case of peak hour demands, this table demonstrates again that average metered demand was somewhat little less than the 50% that of unmetered demand. Also shown in the table, is another basic fact, namely that metering has a much more significant impact on outdoor water use than it does in indoor use. Grima (1972) reflected this fact in his finding that metering has most effect on the less essential water uses. Sprinkling and related use effect the maximum day demand to a much greater extent than indoor use and the peak uses are most relevant to rate design, planning, and financing of municipal water utilities. Thus, metering may reduce the need for storage capacity installed to meet peak demands.

The metering decision has two dimensions, equity and efficiency. The equity argument for metering is clear; it permits volumetric charges, which means payment according to the quantity of water used (and approximately the quantity of waste disposed into the sewers). The efficiency argument can be approached by considering metering like any investment making decision, as characterized by the excess of benefits to costs. Hanke (1981) used a “shortcut” procedure for the benefit cost decision on the extension of metering in Perth, Australia. He found that total benefits during the period of study were just over \$783,000 while costs were just over \$241,000. Thus the excess of benefits over costs was about \$542,000. This finding exemplifies many other studies that have been conducted on metering around the world.

In general, unmetered customers have no incentive to use efficiently; for example they have no incentive to repair indoor water using fixtures, or even to cease lawn watering during rainstorms. Thus, the finding that metered usage par capita is from 30% to 50% lower than that of unmetered customers is not surprising. Table 4.4 shows the results of a number of studies of the impact of metering on water demand and indicated that these impacts are substantial. The implication of this finding is that “universal” metering will have a large impact on system capacity, and therefore required investment.

Table 4.5 summarizes the metering-related results of a recent water use and pricing study in Ontario for the period of 1989-1999.³⁸ The impact of water metering on water use can be deduced from this table. Most of the communities in the two smallest population size categories are dominated by the use of flat rate water pricing practices, and can be inferred that these municipalities have no water meters. The table shows that per capita water usage declines significantly from the smaller municipalities to the

³⁸ This study was carried out to support SBC study #4, being led by PwC.

largest communities in Ontario. Generally, the latter are mostly or fully metered with respect to their water supplies. These communities have significantly smaller per capita residential water usages. This demonstrates again that the impact of water metering on municipal water demand is substantial.

Table 4.4 The Effects of Metering on Municipal Water Use – Selected Studies

Area	Impact and Special Details	Source
Western U.S.	- unmetered areas have over 50% higher water use than metered ones on average; over 100% for maximum day and maximum hour.	Linaweaver, Geyer and Wolff (1967)
Etobicoke - Ontario	- unmetered areas have 45% higher water use than metered areas of comparable assessment.	Grima (1972, p. 165)
St. Catharines, Ontario	- an 11% drop immediately following metering but rebounds because prices kept low. Two years later, water usage higher than before metering.	Pitblado (1967, p. 46)
Alberta	- 10-25% drop in water use following meter installation	Associated Engineering Services Ltd. (1980)
Peterborough, Ontario	- 10% reduction in water use predicted following meter installation.	Peterborough Water Department (1984)
Central Valley, California	- household water use reduced up to 55% following meter installation; usage averaged 30% less in metered than in unmetered cities.	Minton, Murdock and Williams (1979)
Calgary, Alberta	- unmetered water use 46% greater than that in metered residences.	Mitchell (1984)
Calgary, Alberta	- unmetered water use 65% greater than that in metered residences.	Kellow (1970)
Dallas, Texas	- experienced a drop in water demand of 43% following meter installation.	Shipman (1978)
Gothenberg, Sweden	- per capita use in unmetered apartments 50% higher than in a single family, metered residences	Shipman (1978)
York County, Pennsylvania	- substantial increases in industrial waste treatment charges led to reductions in water use in the range.	Sharpe (1980)

Source: Tate, 1990

Table 4.5 Monthly Domestic Water Use per Capita, by Population Size Group

Size Group	('000 liters per capita per month)				
	1989	1991	1994	1996	1999
(('000 persons)					
1 - 5	14.05	13.30	13.20	13.16	13.34
5 - 10	13.93	12.16	12.60	13.03	14.10
10 - 50	12.40	11.93	11.10	10.35	10.65
50 - 100	10.88	10.03	9.60	9.21	9.76
>100	9.23	9.56	9.60	9.49	10.45
Total	10.42	10.43	10.20	9.81	10.54

Source: Data prepared for SBC study #4 by PwC

4.6.2 Non-Revenue Water³⁹

Non-revenue water is defined as water supplied into the municipal water system but is not paid for by any particular customer. This has traditionally been referred to as unaccounted for water and can be calculated by subtracting the volumes accounted for by retail sales from total water supplied into a system. Non-revenue water typically includes water and sewer main flushing, firefighting and training, metered inaccuracies, park watering, water withdrawn from hydrants for municipal functions such as street washing, and water main leakage.

Water and sewer main flushing account for a portion of the total amount of non-revenue water. This quantity of water is necessary in the daily functioning of the water system in order to clean the water and sewer infrastructure.

With respect to fire charges, many communities have a special charge incorporated into their tax systems. It is interesting to note that the first Canadian water systems were built to protect municipal facilities from fires with this function continuing into the present day. Minimum standards for pressure and flow for a specified duration are used to determine firefighting requirements. These guidelines are established by the Insurance Bureau of Canada, based on factors such as the type of building construction, the type of use, the proximity to fire hydrants, and the proximity to adjacent buildings. The requirements for fire flow impose a requirement for additional capacity in excess of normal system demands. Fortin *et al.* (2001) indicate that the capital costs of a water supply system that can be attributable to fire flows may range from 15% of total capacity for larger communities up to 75% for smaller communities, because the fundamental requirement for both is the same. The requirements for fire flow are important also because their impact on system capacity is much higher than their impact on total demand. For example, while the requirement for additional capacity may be 15% of total capacity the impact on actual volumes of water used may only be 1% to 2% of the total.

Water withdrawn for municipal functions also accounts for a portion of the non-revenue water. These tasks include street washing, park watering, etc. Water is directly taken out of the municipal water system through hydrants or different connections to accomplish these jobs and is not controlled. These uses are also not accounted for in the municipal water system billings.

System leakage is a particularly interesting factor in water utilities. In engineering practice an attempt is made to keep leakage within 15% of total water use. However in some systems, which are old or not well maintained leakage rates may exceed 40% of total water supply. This, in effect, means that leakage may be the largest "demand" on some municipal water systems. This is particularly important in smaller communities, with insufficient financial resources to permit adequate system maintenance.⁴⁰

4.6.3 Household Characteristics

Several household characteristics can influence residential water demand. Various studies have included the number of persons residing in the dwelling, the number of water using fixtures, lot size, and lawn and garden size. Many of these variables are multi-linear and co-linear with income, and may be subsumed by the latter. In general, as with the case of income, all of these variables show a direct relationship with water demand.

³⁹ This section draws on the work of Fortin et al. (2001)

⁴⁰ In some underdeveloped countries, one of us has seen rates of leakage up to 60% of total water supply. This represents a tremendous drain on the financial resources of these systems. (WSSCC, 1997).

4.6.4 Urban Structure

Municipal structures vary widely across the province. One effective means of viewing structure is through “central place theory” (Abler, Adams, and Gould, 1971, pages 366 to 378), which holds essentially that the larger a municipality, the greater the number of services (i.e., commercial establishments) and industries located there. In terms of water use, this variable can be examined either through recording the percentage of total water used in the residential, commercial, and industrial categories, or by recording simply the proportion of residential to total use. With respect to the latter, this proportion would fall as population becomes greater.

4.7 Other Independent Variables

4.7.1 Climate

The local climate, particularly precipitation, has a significant, and inverse, effect on water demand. Summer peaking is illustrative. In drier areas summer peaks in water demand are higher than in more humid ones, because outdoor vegetation requires more water, and because of the use of evaporative cooling. In cold climates, there may be winter water demand peaks, as pipes are bled to prevent freezing. Seasonal demand patterns are, therefore, important in planning the size of water supply capacity.

Climatic factors have been dealt with in a number of ways in municipal water demand studies. Foster and Beattie (1979) used the concept of “effective rainfall” (rainfall occurring only when monthly average temperatures were at least 45° F and 60° F in northern and southern U.S. regions respectively). They found that the observed regional variations in estimated own-price elasticity in the two types of water-availability regions were partially due to differences in the ratio of outdoor to indoor water use because these uses have different price elasticities of demand. Howe (1982) used a moisture deficit measure that is a function of outdoor irrigable area, average summer potential evapotranspiration, average summer precipitation rate and the proportion of summer precipitation that reaches the root zone. Counter-intuitively, Howe’s work showed that the moisture deficit coefficient of regression was positive and significant for Eastern summer demand but negative and insignificant for Western summer demand. In other words, the importance of precipitation as one of the determinants of residential water demand is not restricted to arid and semi-arid areas. Hansen (1996) included a “sprinkler needs” variable in his model of residential water demand in Copenhagen (measured by amount of precipitation in summer months). The results of this research indicated that total household demand is an increasing function of sprinkler requirements.

An alternative research strategy has been to use disaggregated data that allow estimation of seasonal demand functions. Sewell and Roueche (1974) used this methodology to examine residential water demands in Victoria, British Columbia. Four dependent variables were used: total annual water demand; peak period demands (June through August); off-peak demands (October through April); and mid-peak demands (May through September). Independent variables were average water prices for a sample of the city’s water customers, income (measured as average disposable income per tax return in different census tracts across the urban region, peak and mid-peak average temperatures, and peak and mid-peak precipitation). The study included all water uses in the municipality, with the assumption that, for Victoria, commercial and industrial water demands were small enough not to affect the primary investigation of residential water demands. Precipitation was found statistically significant in explaining annual and peak demands, but not off-peak and mid-peak demands.

Danielson (1979) used monthly data to estimate residential water demand in North Carolina. The author assumed that water consumption in winter months is due to indoor usage; during the summer it is split between outdoor and indoor use. Outdoor use calculated as the difference between total water use and estimated indoor use, in which the latter is assumed to be constant during winter and summer seasons. Estimated equations indicated that temperature has much stronger influence over summer outdoor demands than winter indoor demands. Moreover, the price elasticity was higher for summer outdoor demand (-1.38) than for winter demands (-.305). Griffin and Chang (1990) estimated a series of residential water demand equations using aggregate community-level data to test whether price elasticities are sensitive to seasonal and/or functional form. They found that residential demand was sensitive to both.

There is a consensus that increases in temperature or evapotranspiration rates generates higher residential water demand, while increases in precipitation has the opposite effect. In addition most of the climate impact is due to summer outdoor uses. Scheffter et al., in an unpublished study, concluded that a 1% rise in (average) temperature would increase residential water demand from .02 to 3.8%, and a 1% decrease in average precipitation would increase it from .02 to 3.2 %. Indoor water use is much less sensitive to climatic conditions.

Areas that use flat rate pricing practices tend to have high partial regression coefficients between income and residential water demand as reflected by maximum day demands. This shows that when the marginal price for water is zero, water users are more sensitive to drought – that is, they are more likely to protect lawns and other outdoor vegetation. In other words, the demand for domestic irrigation water in flat rate areas is more sensitive to climatic conditions (Billings and Jones, 1996).

4.7.2 Water Conservation Programs

Many municipalities in Ontario have instituted water conservation programs. These consist of a variety of measures taken to decrease total and per capita water usage. In broad terms, efforts to change pricing practices within a municipality can be considered as a water conservation measure, although in this paper it has been separated because of its importance for the overall topic of the paper. Other conservation measures may include leak detection and repair, efforts to install low water use fixtures in homes and buildings, public education programs, and many more steps. There is some broad evidence that water conservation programs can have an impact on water use. For example, per capita water use in Canada fell from 10,800 liters per capita-month in 1991 to 9,750 liters per capita-month in 1996 (Table 4.5) . This was attributed by some managers to the effect of aggressive conservation programs. The figures for Ontario (Table 4.5) indicate the same trend, although the corresponding figure had risen to 10,540 liters per capita-month in 1999. Thus, it is difficult to draw any firm conclusions as to the effects of water conservation programs undertaken in municipalities.

4.8 The Demand for Residential Water Services – Theoretical Studies⁴¹

4.8.1 Residential Water Demand

In Ontario, residential water use accounted for approximately 60 percent of total municipal water use on average in 1999. Residential water use is also the use that has been studied most in the academic

⁴¹ This discussion draws very heavily on work by Renzetti, which is currently in the publication process. The author, a member of the GA team, has consented to the use of this material here.

literature. Therefore an overview of some of these studies will be beneficial in illustrating some of the economic concepts discussed throughout this chapter.

The theoretical bases for studying residential water demand from an economic viewpoint is based upon the theory of consumer demand. The mathematics surrounding this theory is fairly complex and will not be presented here. However, one of the main findings is that the quantity of water demand is a function of the price of water, the price of other services and goods, and household income. Despite the fact that these three variables are important in determining water demand from a theoretical point of view, most studies of residential water demand have considered water as separate from other goods and services and do not take prices of latter into account. An additional problem in studying residential water demand arises because there is often no independently determined price for water, and by using quantity used to determine price paid, the customer is simultaneously determining both the level of demand and the marginal price for water. A good example is the case where average price is use as an independent variable in a regression equation to explain the level of water demand. For example, a hypothesis might be set up in which the level of water demand is a function of average price, income, precipitation, and other household characters. Because average price is implicitly a function of the total amount of water used, the demand for water appears on both the right hand and the left hand side of the regression equation, and therefore the level water demand and the every price are determined simultaneously. This is called a simultaneity problem, and has been the subject of many econometric studies. It will be discussed in more detail below.

A large number of studies have been devoted to determining the relationship between residential water use and a number of explanatory variables suggested by economic theory. Throughout these studies, a number of problem "dimensions" have been examined. These dimensions include the appropriate definition of price, the choice of estimation procedures, and the role of other non-economic explanatory variable, especially climate.

Before proceeding with a discussion of these various dimensions, it is beneficial to point out a number of general challenges. First water demand is a complex function of quantities used. For example, many municipalities employ either increasing or decreasing block rates to determine the amount of money that a consumer will pay for water services. In addition there may be sewer charges added to of these block-rate-determined prices. Therefore, researchers must deal with often very complex pricing structures in the determination of price. Second, often variable quantities of water are included free of charge in the rate schedule (i.e., the marginal price of this water is 0). The justification for this procedure is to insure that lower income groups are not deprived water services. In the terms of Chapter Two, this is one reflection of the concept of equity, which may or may not be beneficial in trying to establish effective water rates. Third, in high income countries, water is generally of a relatively small item in the budget of most consumers. For example, in Ontario for 1999, the average price paid for water services by residential customers averaged just over \$33 per month (at the 25 cubic meters per month level), as shown in Table 6.2. While Dinar and Subramanian (1997) have pointed and that this is low by world standards, nevertheless it is the case that pertains to Ontario. This leads to the question of whether consumers are even aware of the amount of money being spent on water? If the answer to this basic question is negative, the study of the role of price in determining residential water demand is made exceedingly difficult. Fourth, in many cases data are unavailable on potentially important variables such as the stock of household water using capital, and the precise characteristics of potable water requirements (pressure, reliability etc.). Finally there are a number of econometric issues that also pose difficulties -- principles among these are the selection of the functional form of water demand equations, the definition of explanatory variables, and the selection of estimation procedures.

a. The Influence of Water Prices

The first modern studying of Residential water demand was by Howe and Linaweaver (1967). This study was based on a research project on residential water demand carried out at Johns Hopkins University. The object of the Howe and Linaweaver study was to calibrate regression equations to explain the

demand for indoor (domestic) and outdoor (sprinkling) water use. The authors were also interested in the effect of rate structures and metering on water demand, and the observations were further differentiated by region in order to explore the effect of water availability conditions. The data used were average values for each observation, in this case individual municipalities. The dependent variable under discussion was average water use per dwelling per day. Outdoor water use was taken as summer use minus average winter use (which was considered to be indoor use). Prices were determined as the sum of the marginal water and sewer blocks into which the average water use fell. Property value was used as proxy for income, with other explanatory variables including the age of the dwelling units and the number of residents per household. The study showed that domestic water use was inelastic with respect to price (0.21 to -0.23). Sprinkling demands were more elastic, with a range of price elasticities from -.44 to -1.57. This shows that sprinkling water demands may fall into the elastic range with respect to price. Income elasticities showed essentially the same results: the income elasticity of demand for indoor water use was from 0.31 to 0.38, and for outdoor use 0.45 to 1.45.

Howe and Linaweaver identified the simultaneity problem in their study. As already noted, this source of possible bias is related to the specification of price of water; if the price is exogenously determined and constant (as it is for most marketed goods and services), the estimation of water demand regression equation is straightforward. But price is often not constant (Dinar and Subramanian, 1997), and therefore the user determines the marginal price paid in selecting the level usage, thereby introducing the simultaneity problem. One solution to this problem has been to employ average cost of water as proxy for price. Foster and Beattie (1979) used this approach to estimate an aggregate water demand equation. In this study, the dependent variable was average water use for household per year, with the independent variables being average price of water per municipality, median income, the number residence in each household and dummy variables to represent geographical regions. This study produced price elasticities of demand ranging from -0.3 (Midwest) to -0.69 (Pacific Northwest). The authors acknowledged the simultaneity problem. Griffin and Martin (1981) criticized the study for using average cost to estimate marginal price and claimed that this was an improper identification of the demand model.

Economic theory, as indicated in Chapter Three, specifies that marginal price is the appropriate variable can be use in measuring price of water, rather than average price. However, using marginal price produces certain problems. First, as shown above, marginal price may not be constant, implying that it is co-determined with the volume water use, and therefore that the resulting regression equation is probably misidentified and contains a simultaneity bias. An additional simultaneity bias may rise from measurement error. For example, the marginal price paid by each household will not be apparent in aggregate observations on water demand are used.

Billings and Agthe (1980a) computed a single demand equation for Tucson AZ, using a time series of monthly data. The city used an increasing block rate to calculate residential water bills. In this study, the dependent variable was the average water use per household. The independent variables included the marginal price at average level of demand. Griffin and Martin (1981) criticized the use of marginal price with aggregate data introduces possible bias into the coefficients because errors in measurement may mean that the marginal price (and thus the "difference variable" [defined below] may be correlated with the error term. The argument can be stated as follows: when the variance of error term small, most observations will be close to the demand block where demand curve intersects price schedule, and, accordingly, the regression line will be close to "true" demand curve. But, if the variance of error term is larger, an increasing number of observations will fall into the inner and outer blocks where different marginal prices will apply.

Nauges and Thomas (2000) recently identified an even broader source of simultaneity. They sought to estimate price elasticity for residential water demand taking all sources of simultaneity bias into account. In addition to pricing simultaneity, they claimed that utility managers take community characteristics (population density, average income, etc.) into account when setting price. These factors

influence water use, thereby producing a further form of simultaneity bias. To examine this bias in more detail, Nauges and Thomas first estimated a price equation to identify the factors influencing rate schedule choice. Then, they estimate a water demand equation in which the dependent variable was the average annual residential demand, with independent variables consisting of price, income, climate, and a vector of community characteristic variables such as population density, the age of housing stock, the proportion of households metered, and others. Data comprised aggregate observations from 116 French communities over the 1988-93 period. For the demand equation, income and the proportion of new housing are significant in explaining choice of rate schedule. For the estimated demand equation, "Hausman specification tests indicate exogeneity of the explanatory variables, and among them the average price of water and the presence of local community effects."⁴² This indicates a possible alternative way in which endogeneity can appear in a demand equation.

A second problem with the use of marginal price is that even if two identical households face same marginal price they may still differ in their consumption because of other features in the price schedule. Taylor (1975) pointed out that if one household faces a declining block rate schedule, while another an increasing block rate, their total water-related expenditures differ and thus will their residual incomes. These differences in residual income may have an impact on water demands. Taylor suggested including both average and marginal prices in the demand equation to account for differences in intramarginal price blocks. Nordin (1976) proposed an alternative procedure – inclusion of a "difference variable" (D) in the demand equation. This difference variable would be equal to the household's actual water bill minus what it would have been had all units of water demand been charged at the marginal price.

A number of authors (Howe, 1982; Saleth and Dinar, 1997) have argued that this difference variable would not reflect all the characteristics of intramarginal price blocks and that it may be a separate source of simultaneity bias (because it is a function of marginal prices). However, a number of papers have adopted the Taylor–Nordin approach to specifying prices, and employ marginal prices and D in their estimation models (Billings and Agthe, 1980a,b; Polzin 1984; Nieswiadomy and Molina, 1989; Saleth & Dinar, 2000). The results have been mixed at best. Billings and Agthe showed that D should have an estimated coefficient equal to but opposite in sign from the income variable. A number of studies have found that the difference variable has a negative and significant coefficient, but in most cases the size of the resulting coefficient has differed from the income coefficient by an order of magnitude (e.g. Billings and Agthe, 1980a; Jones and Morris, 1984).

An alternative to the Taylor-Nordin price specification approach and OLS is to adopt an estimation procedure that addresses the simultaneity problem specifically. This is done by creating an instrumental variable to serve as a proxy for price, or by using a two-stage regression procedure in which price and quantity are determined simultaneously. Jones and Morris (1984) provided an example of first approach. They pointed out first that when quantity (and therefore price) are measured with error, even using the 'correct' price can produce a correlation between the price variable and the error term, and that this will produce biased estimates. Their solution estimates an instrumental variable⁴³ for each of the price specifications (average, marginal and D. For the marginal price variable, the following is used: "Average summer and winter water use are computed for each rate class of the sample. A summer or winter marginal price is associated with an estimate of typical water use in that season through the rate schedule relation." The instrument for the D then uses marginal price. The average price instrument is created by regressing the average price for each household against features of the rate structure (which are exogenous to the decision maker). For the double log specification, price elasticities for marginal price, average price and D are -0.21, -0.34, and -0.23 respectively. Income elasticities are 0.40 and 0.46 depending on the price definition. This approach has been used a number of times (Renzetti 1992a; Nauges and Thomas (2000).

⁴² Quoted by Renzetti, 2001

⁴³ An instrumental variable is highly correlated to the variable of interest, but orthogonal to all other explanatory variables.

An alternative to the instrumental approach is a 2-stage regression model. Chicoine, Deller, and Ramamurthy (1986) calculated a series of equations in which demand, MP, AP, and D were estimated simultaneously. The equation contains a third price (in addition to MP, D, income, the number of residents per household and the number of bathrooms) and this is AP-MP (defined as OP). Opaluch (1982) had shown that one can include OP to test whether consumers respond to MP or OP. The authors used monthly consumption data for Illinois rural households plus all households facing a declining block rate structure. They used three models: a single equation OLS, a two stage least squares (2SLS) and a 3SLS. The analysis produced several interesting results: in no case did D meet expectations (i.e., equal in magnitude but opposite in sign to the income coefficient). Thus they conclude that the Taylor-Nordin specification may not always be the best description of consumer behaviour; second, the coefficient estimates were roughly the same across the 3 models, although the 3SLS appears to produce more efficient estimates. Estimated price elasticity with respect to MP is -0.22 (OLS), -0.42 (2SLS) and -0.42 (3SLS). The income coefficient was only significant in the 3SLS mode (.14). In a related study, Nieswiadomy and Molina (1989) examined OLS bias under declining and increasing block rate structures. They compared OLS, 2SLS and instrumental variable estimators of a single equation demand model. Data were based on a set of monthly, household-level observations over two time periods - one in which the households faced DBRs and the other where they faced IBRs. These authors used a Hausmann specification test to show strong evidence of bias in OLS estimators. The models also showed that D does not behave as the Taylor-Nordin formulation predicts. The price elasticity ranged from -0.36 to -0.86 across the IV and the 2SLS estimators.

While the two stage regression procedures outlined in the last paragraph are accepted methods of dealing with the endogeneity issue, they do not directly address which price specification consumers actually use. Marginal cost pricing is suggested by economic theory, but this assumes that the cost of acquiring information is zero. If information is costly, another price specification might be more appropriated. This problem was first addressed by Opaluch (1982), then by Shin (1985). The latter, studying electricity costs, found strong support that customers used ex post average prices as the basis of their perception of price.

Nieswiadomy (1992) considered this problem of correct price specification for residential water by conducting three regression analyses: one using marginal prices, one using average price, and the third using Shin's price perception model. The data used were taken from an AWWA cross-sectional survey of water prices and demand for major U.S. cities, divided into regions. Nieswiadomy also used dummy variables to represent the presence or absence of conservation or public education programs. In this study the Hausman test for endogeneity failed, meaning that the coefficient estimates were biased. Also, the estimation of the price elasticities was small (-0.29 to -0.45). Also, the income elasticities were small (0.14 to 0.28). The estimates of the Shin price perception variable supports the hypothesis that consumers rely on average price for their perception of water price. Conservation programs, in this model, had no significant effect on water demand, but public education programs did reduce demand in a statistically significant manner, particularly in the South.

A recent development in estimating residential water demand concerns alternative ways to test the non-linear nature of the budget sets that occur under decreasing and increasing block rates. Hewett and Hanneman (1995) employed a discrete/continuous choice model. This model assumes that customers employ a two-stage optimization process in selecting their level of water demand. The first stage is the decision about which blocks of the pricing schedule in which consumption should occur, and then deciding quantity to consume within the block. This model produced price elasticities much higher than those obtained from most studies, in the rate of -1.5 to -1.7, although the income elasticities were more in normal range (0.15). Although the results are interesting, it seems unlikely that water customers make their decision in this manner, and therefore it is felt that these results are unusual.

More recently still, Pint (1999) estimated both a fixed effects and a Hewlett-Hanneman model to investigate the responses of California household to significant price increases during the state wide prolonged drought. This drought produced substantial price increases, and the demand equations could

be estimated using much wider price dispersions that had been the case in previous studies. The independent variables in the model included water price, water price squared, house and lot size, and a variable for climate. Interestingly, the findings showed that the ordinary least squares model and the fixed effects model have positive slopes at higher price levels. Estimated price elasticities were in the range of -0.20 to -0.47 (Summer) and -0.33 to -1.24 (Winter) for the Hewett Hanneman model. These results have not been replicated, and therefore should be treated as hypothetical only. This is one of the few studies to find that winter price elasticities are higher than summer ones.

The question of functional form use for demand estimation has received less attention than the choice of estimate technique and the estimation of price variable. Prior to the mid-'80s, several linear forms had been used in demand estimation. For example, linear, log-linear, and double log functions had been used. Many studies, however, had failed to draw the link between the selection of the functional form of the regression equations used and the underlying theory. (One good example the contrary can be found in Grima (1972).) Functional form has direct implications for estimation results -- particularly on estimated price elasticities. For example the price elasticity is constant in a double log model, whereas it is variable in a linear model. Choice is often ex post, that is after plotting the data prior to estimate. In other words, the form selected might simply be the form which appears fit the data best, which produces the greatest degree of explanation, with no link at all to underlying theory. This strictly empirical approach offers little in the way of theoretical explanation.

Billings and Agthe (1980a) experimented with this problem estimating both linear and double log model. They found that the price elasticity obtained using a double log model was -0.27 , while under the linear model, the elasticity ranged between -0.45 and -0.61 . This shows that the selection of functional form can be a factor in analyzing the price elasticity of water demand.

There have been a limited number of studies that have experimented with more complex functional forms. For example, Al-Qunaibet and Johnston (1985) employed a linear expenditure system model to estimate residential water demand for Kuwait. This functional form had the advantages of being consistent with maximizing behaviour, having expenditure on each good being a linear function of prices, and being able to determine endogenously the 'subsistence' level of consumption for each good. Despite these advantages, few water demand papers appear to have employed the flexible functional forms that have been used extensively in other fields of demand research (Pollak and Wales, 1992).

The types of databases used for residential water demand estimation have evolved over the past two decades. Early demand studies relied upon single period, cross sectional, aggregate observations (Saleth and Dinar, 1997). Over time, a number of time series and pooled databases have begun to be used (Agthe and Billings, 1980; Carver and Boland, 1980). Household level databases containing information on consumption, prices, and household characteristics, are becoming more common. Renwick and Archibald (1998) employ household level observations on water use, prices adoption of conservation strategies, and household characteristics in studying residential water demand

The development of micro-level pooled databases has paralleled use of more sophisticated techniques of estimation. It was shown earlier that Pint (1999) estimated a fixed effects model and a discrete/continuous model using a dataset of household level observations. However, it is not yet clear whether the type of database used influences the price and income elasticity estimates resulting. A once-common view held that cross-sectional databases reflect long run trends better than time series data and could be expected to produce larger price elasticity estimates (Boland et al. 1984). However, based on a meta-analysis of the residential water demand literature, Espey et al (1997) concluded the price elasticities produced by the two types of study are not statistically different.

The discussion to this point shows that a variety of price specifications, model specifications, data types and estimation procedures have been used to estimate residential water demand. As already pointed out, one of main aims of these studies has been to determine sensitivity of residential water demand to the price, and changes in the price, of water. It is logical to ask whether the wide array of

possible choices has made a discernable impact on resulting estimates of price elasticities. Espey et al. addressed this issue by conducting a meta-analysis of econometric residential water demand studies, whose purpose was to identify factors that influence elasticity. These authors used the empirical findings from 24 journal articles (some with >1 estimate of price elasticity). The analysis regressed the estimated elasticity against a set of features implicit in the data or the estimation model drawn from the articles. The independent variables were a set of binary variables indicating the presence or absence of any one particular factor in the particular publication. The factors were 1) demand specification – whether specific factors like income, population density, climate, season or lagged dependent variable were used; 2) data characteristics – household level or aggregate, time series, cross sectional or pooled, and type of price (MP, AP, Shin, or difference; 3) the environmental setting – western or eastern US; and 4) the type of estimation technique used – OLS or some other technique. For the DV (i.e., price elasticity) short run values ranged from -0.33 to -2.23 (median -0.38); long run values ranged from -0.1 to -3.33 (median -0.64). The following features were found to be significantly and positively correlated with the price elasticity estimates: long-run specification, pooled residential-commercial data sets, summer demands, a price specification other than marginal, and the use of increasing block rate schedules. In contrast, the inclusion of rainfall as an independent variable was significantly and negatively correlated with price elasticity. Insignificant factors in explaining elasticity included: location, population density, household size, temperature, income, cross-sectional vs time series data, household level vs time series data, consumption interval, functional form and estimation technique. These are very interesting findings, and can be used as guides for future research.

b. Influence of Other Prices

Economic theory predicts that water demand is a function of both its own price and all other prices faced by a household. In particular, we would expect the prices of sewage treatment, energy, and water related capital could each influence residential water demand. Relatively little attention paid to influence of other prices in the published literature. Renzetti claims that this is in part because a number of researchers normalize water prices with an index of consumer prices that is taken as a proxy for all other prices.

There are some studies that include other prices. Hansen's (1996) model of residential water demand includes electricity price as one independent variable. Hansen used time series data for average household water use in Copenhagen. This study found own-price elasticities that are very small and insignificant, but estimates cross-price elasticities with respect to energy to be -0.21.

Another reason for ignoring other prices in water demand equation arises from the practice of defining the price of potable water to include sewer charges (Howe and Linaweaver, 1967). There are exceptions to this practice. Billings and Agthe (1980a) included a sewer price as a separate independent variable, but dropped it when it proved insignificant. Renzetti (1999) included prices of electricity and sewage treatment in a model of residential water demand. Data in this study were aggregated, annual, cross-sectional observations from a sample on Ontario municipalities. Estimates of elasticities of residential water demand with respect to sewer charges and electricity prices are -0.16 and -0.284 respectively.

c. Summary

Considerable effort has been expended to characterize the structure of residential water demands. Over the last thirty years, part of that effort has been directed to employing increasingly sophisticated model specifications and estimation techniques. In addition, a central concern has been the specification of the price of water when consumers are faced with complex rate schedules. The results of these efforts, for the most part, are fairly clear. Economic factors such as the price of water and household income clearly play an important role in determining residential water use. Nonetheless, it is also quite clear

that residential water demands (with the possible exception of outdoor water use in summer months) are price and income inelastic.

One of the less clear results, however, concerns the specification of the price of water. Despite the attractiveness of the Taylor-Nordin specification on theoretical grounds, its performance has been disappointing. It may be, as a number of authors have asserted recently, that the billing formats and informational constraints facing consumers mean that there is no single 'correct' price of water. Rather, the actual price used by households in their water use decisions is influenced by the presentation of rate schedule information and, thus, must be investigated on a case-by-case basis. It also appears that the attention paid to the specification of the price of water has meant that other features of residential water demands have not yet been explored fully. These include the role of other prices, the possibility of estimating daily or even hourly water demands and the role of characteristics of water (such as reliability and quality) in determining household demands.

4.8.2 Demand for Waste Treatment Services

This section examines what is known regarding households' and firms' demand for wastewater disposal. Of particular interest is the extent to which research has demonstrated that water disposal decisions are sensitive to market conditions and economic instruments. While there has been a substantial amount of research conducted by engineers and scientists concerning the chemical and biological features of wastewater and its treatment (cf. the journals *Waste Management*, *Water Resources Research* and others), there has been relatively little attention paid by economists.

Almost any consumptive use of water implies an alteration in its features. These alterations include changes to water's chemical composition, microbiological character, temperature and clarity. As well, changes may occur in the levels of contaminants such as suspended solids, bio-oxygen demand (BOD), fertilizers, pesticides, metals and chlorine or nitrogen-based compounds that are present in the discharge water. Furthermore, unless all of the water withdrawn from the environment is consumed (for example, included in a firm's final output or lost through evaporation), then any water intake also implies a need to dispose of some of that water. Disposal can take several forms including physical connection to an off-site sewage treatment facility, on-site lagoons, evaporation, direct release into a surface water body, injection into an underground water body or well or, in the case of agriculture, absorption by soil and surface-run-off.

The discussion here proceeds along sectoral lines and begins by examining households' sewage disposal demands. Largely because of the lack of metering and pricing of residential sewage flows, there has been very little research directed at studying the economic features of households' production of wastewater. Industrial demand for water disposal differs significantly from that of households. Firms may have a variety of options available to them in their waste disposal that are not typically available to households. Furthermore, firms are more likely to be subject to direct environmental regulations regarding the characteristics of their wastewater flows. In fact, some observers suggest that increasingly stringent effluent regulations are a major factor in explaining the reduction in water intake observed in North American manufacturing firms (Solley, Pierce and Perlman, 1999).

a. Residential Demand

The quantity and quality of wastewater flowing from a household is related to the flow of water into the household and the household's use of that water. However, the flow of water from a household to the

municipal sewage collection system is only imperfectly correlated with the flow of water entering the household. On the one hand, it is reasonable to assume that almost all of the water used for indoor purposes will go directly to sewers. On the other hand, not all water applied to outdoor uses (gardening, car washing, etc.) returns to the sewage collection system. In addition, in the case of combined systems that collect wastewater as well as surface run-off, not all of the water entering the collection system emanates from firms and households. As a result of these factors, there is not a perfect correlation between aggregate water intake and the total flows to sewage treatment facilities (Tchobanoglous and Schroeder, 1987).

There are substantial obstacles standing in the way of modeling households' demands for wastewater disposal. First, it is very rare for household sewage flows to be metered. Second, residential sewage pricing is much rarer than water pricing (Dinar and Subramanian, 1997). The financing of sewage treatment services is usually based on pricing formulas that have little, if any, connection to actual household wastewater characteristics or flow rates. The charge for sewage treatment is commonly based in North America on property value or property frontage. To the limited extent that households face a nonzero marginal price for sewage flows, it is common practice for municipalities to define the price of sewage disposal as a percentage of the price of potable water.

Another complicating factor confounding the modeling and estimation of household demands for water disposal is the question of whether there is evidence that households even consider their generation of wastewater (Cameron and Wright, 1990). In addition, even if it assumed that households do decide on the quantity of wastewater to emit and the types and quantities of contaminants to place in that flow, there still exists the problem of modeling the demand for sewage treatment. Specifically, it is not clear what households would be demanding – simply the physical removal of wastewater or removal and some level of treatment. That is, do households have an expectation regarding the level of treatment that is carried out and is that expectation the same or different from what is actually carried out?

The most common approach to modeling the determinants of residential wastewater flows has been to assume that they are primarily determined by population, the features of stock of water-using residential capital (sinks, showers, toilets, washing machines, etc.) and weather (Tchobanoglous and Schroeder, 1987). For example, Billings and Jones (1996) indicated that the per capita water use for toilets and showers is in the range of 10-15 and 10-20 gallons daily. Under this approach, changes in residential wastewater flows arise primarily from changes in weather patterns, residential population and changes to the stock of capital (e.g., the installation of low-flow toilets). Changes in residents' water use habits (due, for example, to education or conservation programs) are also a possible source of changes in wastewater flows. Under this approach, relatively little attention is paid to the possibility that economic incentives might alter behaviour except for the possibility that subsidizing the installation and retro-fitting of water-conserving capital may have an influence (Cameron and Wright, 1990; Renwick and Archibald, 1998).

Renzetti (1999) conducted a study to estimate the relationship between municipal wastewater flows and economic factors. The broader purpose of this research was to study and evaluate the pricing practices of municipal water supply and sewage treatment utilities in Ontario. Water and sewage utilities' estimated cost parameters were combined with estimated residential and non-residential demands for potable water as well as an estimated aggregate demand for sewage treatment function in order to calculate approximate welfare losses arising from under-pricing and over-consumption of water.

An aggregate demand equation was estimated for sewage treatment services, because the utilities in the sample did not record residential and non-residential flows separately. Aggregate demand for sewage treatment (Q_s) was assumed to be a function of the price of sewage treatment (P_s), the price of water supply (P_w), the price of electricity (P_e), average household income (I), number of households (NH), number of firms (NF) and a vector of climate-related variables (V). For each of the three prices in the sewage treatment demand equation, a weighted average of the residential and non-residential

marginal prices was created using each user group's water consumption as weights. The general form of the aggregate demand for sewage treatment equation was the following:

$$Q_s = D_s(P_s, P_w, P_e, I, NH, NF, V)$$

There were 77 aggregate observations used to estimate the sewage treatment demand equation. The equation was estimated using OLS with all the variables expressed in natural logs. In addition, homogeneity of degree zero in prices was imposed on the estimation coefficients and White's correction for heteroscedastic errors was employed. Renzetti reported that the demand elasticities with respect to the marginal price of sewage treatment and number of households were -0.33 and 1.178, respectively. The elasticities of aggregate sewage demand with respect to the prices of potable water and electricity were -0.30 and 0.064, respectively, but neither was statistically significant.

A separate set of studies examined the determinants of households' valuation of improved sanitation services in low-income countries. Whittington et al., (1993) and Altaf and Hughes (1994) conducted contingent valuation surveys in order to estimate households' willingness to pay (WTP) for improved services in Ghana and Burkina Faso, respectively. While both studies' surveys followed the standard format of presenting respondents with alternatives and then eliciting their willingness to pay for them, the surveys differed in how the alternatives are presented. Whittington et. al followed the usual method of describing alternative technologies (for example, water closet with piped connection or ventilated improved pit latrine). In contrast, Altaf and Hughes chose to present the alternatives by describing their characteristics or attributes (for example, whether it required connection to piped water supply and the level of maintenance required).

Both research teams found that WTP for improvements to sanitation service were in the same order of magnitude as households' WTP for water service. Altaf and Hughes, for example, found that households' WTP for improved sanitation ranges from 1% to 4% of total monthly expenditures and from 20% to 87% of monthly water expenditures. Whittington et. al found that the principal determinants of households' valuation of alternative sanitation technologies were income (positively related), current expenditures on water or sanitation (positively related) and the level of satisfaction with existing facilities (negatively related). Both papers concluded by demonstrating that the estimated households' values for improved sanitation can be used to assess the financial viability of alternative infrastructure projects. In the case of households in Ouagadougou, the more costly option of off-site disposal was found to be financially infeasible (households' WTP implying a pay-back period of 20 years), while the less expensive option of on-site disposal was more attractive (households' WTP implying a pay-back period of only 4-5 years).

Despite the limited number of studies directly related to characterizing residential demands for sewage treatment, some indirect evidence can also be gleaned from studies of residential demands for potable water. Because of the complementary relationship between the demands for potable water and sewage treatment, factors that are known to increase the demand for potable water can be expected to also increase the demand for sewage treatment. This relationship, however, is complicated by the importance of information regarding the composition of household effluents. For example, an increase in household income may increase the flow of water entering (and, thus, exiting) a household but it may also change the household's consumption patterns and this may have an influence on the composition of its water-borne contaminants. An increase in income may lead to an increase in the opportunity cost of time and, as a result, members of the household cease changing their automobile's oil at home. This may be important in the aggregate as automobile oil entering sewage systems is a significant challenge to sewage treatment facilities.

b. Industrial Demand

There are several streams of literature that provide information on firms' decisions regarding the disposal of their wastewater. The first is a number of studies commissioned by Resources for the Future in the 1960's. The second is a series of papers written by Russell Thompson of University of Houston and his colleagues in the 1970's. The third is a small number of econometric studies that are concerned with characterizing firms' water-using technology. The fourth is a set of papers whose purpose is to assess governments' efforts to alter industrial wastewater flows through the imposition of charges. Each stream is reviewed briefly.

The work by Löf and Kneese (1968) is representative of the work done by a group of natural resource economists for Resources for the Future in the 1960's (other studies are Bower, 1966; Kneese and Bower, 1968; Russell, 1973). As Löf and Kneese indicated in the preface to their book, the goal of these projects was "to help clarify the impact of the water environment upon industry and to analyze the techniques and costs of changing the impact of industry on the water environment" (p. v). These authors employed detailed engineering process models to characterize the generation of waste products (suspended solids and BOD) in beet sugar processing factories and to investigate the costs of removing these effluents from wastewater streams. A linear programming model was used to depict the firms' technology. The primary output of the authors' work was an estimated function that relates the marginal cost of pollution abatement to the level of effluents contained in wastewater flows. The authors predicted that relatively low-cost adjustments would be undertaken first as the means of reducing effluent flows. Higher levels of reductions in effluent flows require increasingly expensive adjustments. Thus, the marginal abatement cost would increase with the percentage of BOD removed. This finding was consistent with Kneese and Bower's (1968) discussion of how firms respond to the imposition of sewage surcharges:

"The responses of industrial operations to the imposition of sewer charges can be generalized as follows: First, the imposition of a charge or surcharge tends to encourage plants to make changes that in many cases reduce not only the volume of effluents and the wastes in effluents but also the water intake. Second, sewer charges tend to induce an examination of production processes that often uncovers relatively simple modifications that may result in net reduction in total production costs."
(Kneese and Bower, 1968, p. 170).

The second line of research derives from the work of Thompson and his colleagues. In a series of papers published in *Water Resources Research*, these researchers employed linear programming models to describe the water use and disposal processes inside large American manufacturing facilities. For example, Calloway, Schwartz and Thompson (1974) developed a detailed linear programming model to examine water intake and wastewater production for a representative ammonia plant. The authors indicated that part of the motivation for the study was the *Water Pollution Control Act* (1972) that had an expressed goal of zero discharges of certain effluents by 1985. As a result, the authors were motivated to model the impact of a zero-discharge rule on plant operations and costs. The programming model demonstrated that the plant's water use and wastewater were sensitive to prices as was the production of effluents (mostly suspended and dissolved solids). Like many linear programming models, the response of water intake and discharge to charges was discontinuous. At very low prices, once-through cooling is optimal. At a price of 3.2 cents per 1000 gallons, the plant switched to recycling its cooling water and this reduces water intake by 95%. Similarly, an effluent tax of 3.6 cents per pound of dissolved solids led to a 95% reduction in solids discharge and a 3% increase in total plant costs. Further price increases do little to encourage further water conservation and reductions in discharges.

The third set of papers share the characteristic of applying econometric models to characterize firms' decisions regarding wastewater flows. Sims' (1979) study was an effort to determine the responsiveness of firms to municipal 'extra-strength' sewer surcharges. To do this, the author constructed a linear programming representation of the production function for brewing firms, including emissions of BOD and suspended solids as inputs. The dual to this production function model was a cost function assumed to be weakly separable in emissions and water intake from other inputs. The water-related

sub-cost function was then estimated and the estimated unit cost was substituted into the aggregate cost function as a 'quality-adjusted' price of water. Data were pooled cross-sectional, times-series observations from a small number of breweries in London, Ontario.

Sims' estimation model yielded several interesting results. The own-price elasticities for BOD, suspended solids and intake were -0.573, -0.450 and 0.945, respectively, where these elasticity values are somewhat higher than the typical values of residential models. Sims' model was also able to calculate cross-price elasticities. For example, the elasticity of water intake with respect to the sewer surcharge is in the range of 0.044 to 0.062. Thus, it appears that increasing the level of the charge related to BOD and SS emissions actually induced firms to increase their water intake by a small amount. This probably occurred because the municipal sewer surcharge is based, in part, on the concentration of the breweries' emissions. As a result, firms may have had an incentive to use additional intake water not for production purposes but rather to dilute their effluent streams.

Renzetti (1992a) also estimated an econometric model that examined firms' intake and discharge decisions. This paper was discussed earlier, and only the portion related to discharge decisions will be presented here. This study of water used in Canadian manufacturing found that water discharge in some industrial sectors was sensitive to the marginal cost of discharge. For the manufacturing sector as a whole, the own price elasticity of water discharge is -0.9752. In general, the largest water-using sectors display the largest elasticities for water discharges. These include chemicals and petroleum (-0.9302) and paper and wood products (-0.9471). A second finding was that the cross-price elasticity between water intake and discharge was negative (-0.2244) for the pooled manufacturing dataset, indicating the intake and discharge are complements. Thus, if there is an increase in intake prices, it is predicted that both intake and discharge quantities would fall. However, since these data did not contain any information on the make-up of the effluents in the discharge water, it was not clear whether a decrease in discharge flows implied any change in the absolute quantity of contaminants. The other relationship that is of interest was that between water recirculation and water discharge. The data set yielded a range of estimates of the cross-price elasticity between water discharge and water recirculation of 0.0605 to 0.5554. As Renzetti concluded, "This result points to the potential for using economic incentives to reduce industrial pollution. If effluent taxes were imposed on Canadian manufacturing firms, then these firms could be expected not only to reduce the quantity of their water discharged but also to increase the amount of in-plant recirculation."

The final line of research related to industrial wastewater discharges examines the extent to which government policies are effective at regulating those discharges. Despite the extensive efforts by governments in North America and Europe to regulate industrial effluents, there are few studies that document firms' responsiveness to specific features of environmental regulations. Tietenberg (1985) surveyed a number of efforts to measure the cost effectiveness of traditional 'command and control' water quality programs in the United States. One finding of these studies was that the programs achieved water quality improvements at a level of costs well above what could have been achieved under a policy orientated at exploiting firm's differences in treatment costs.

Merrett (2000) pointed out that part of the problem in assessing the impact of effluent charges is the fact that many of these types of charges have been introduced to fund government environmental initiatives rather than to induce changes in behaviour on the part of polluters. An additional source of complication has been that many industrial effluent flows contain a variety of pollutants. A charge directed at one pollutant may induce process changes that lead to reductions in one or more pollutants. Brown and Johnson (1984) described and assessed Germany's 1976 Effluent Charge Law. Under this law, the national government established a series of permits that detailed the allowable effluent levels and a set of charges based on effluent quantities and the quality of the receiving waters. On the whole, the authors were critical of the actual charge levels, stating that:

"The actual effluent charge system bears little resemblance to an idealized one." and

“A charge of 12 DM in 1981 rising to 40 DM per damage unit in 1986 was, and is, too small to achieve the desired water quality objectives for the country and it cannot be a very great incentive to discover low cost abatement technologies.”

Furthermore, they concluded that because all polluters face the same standards independent of abatement costs, the resulting distribution of abatement could not be efficient. Despite these criticisms, the authors noted that there were still a number of positive features of the charges, including the apparent increase in investment in abatement equipment observed at municipal sewage treatment facilities and in some industries.

In another study, Stephan (1988) developed a multi-region, dynamic equilibrium model to examine the impact of water quality standards in Western Europe. Stephan's model contained a lagged adjustment to standards, by imposing the requirement that capital cannot be adjusted instantaneously. Output was hypothesized to be a function of conventional inputs and wastewater, and the two groups of inputs were combined in a constant elasticity of substitution (CES) production function. The data were assembled from parameter estimates representative of the three regions (Northern Europe, Southern Europe and USA/Japan). Using his model, the author compared a base-case (BOD reduction of 60%) to a simulated alternative (regulation requiring BOD reduction of 95%). Total GNP across the three regions declined very little when the more severe regulation was put in place but there were significant differences across the regions. The latter resulted from differing assumptions regarding each region's technology and its ability to adjust to new standards. Under the more stringent regulation, wastewater emissions fell by approximately 25-35%, depending on the region) but these decreases required a number of years to occur, due to lags in capital investments. Unfortunately, despite the fact that the model was based upon optimizing behaviour (firms and households maximize the present value of profits and utility, respectively), the author did not calculate the relative costs and benefits of the alternative regulations. Nonetheless, the author concluded that, “In contrast to most theoretical studies...emission standards provide a dynamic incentive to reduce wastewater generation”.

Finally, Lanoie et al. (1998) employed econometric models to discern whether and to what extent environmental regulation influenced firms' investment, output and emissions (suspended solids and BOD). Data were firm-level observations of Ontario pulp and paper plants operating between 1985 and 1989. The authors employed several econometric models including one that applied an instrumental variables approach to account for the possibility that regulations were not exogenous (in the sense of being based only on the governments' perceptions of the costs and benefits of regulation). In general, the estimation models yielded mixed results. In the case of BOD emissions, firms responded to new regulations by increasing investment, but this failed to translate into reduced emissions. For suspended solids, however, more stringent limits led to reduced output and reduced emissions. These results may be due to two factors: the relatively weak regulations in Ontario (they are less stringent than USEPA regulations); and data limitations that the authors acknowledged (e.g., output was proxied by the quantity of intake water and investment was self-reported).

4.9 Water Demand Management

The results outlined in section 4.8 demonstrate that economic conditions do have an impact on water demand levels. This suggests, in turn, that these conditions, and more broadly, several other socio-economic factors could be used to influence the levels of water demand (and, by implication, financial requirements).

As illustrated in Chapter Three, the traditional approach to managing water has focused on developing adequate supplies of water to meet all uses when they have been required. This has usually been done without consideration of the possibilities of altering pattern and quantities of use – in other words,

viewing water uses as “demands”, in the sense that they can be altered, not as requirements that must be met under any circumstance. The alternative, and much more recent philosophy of management has been termed “water demand management.” It is appropriate at this point of the report to discuss this approach to suggest an alternative framework within which to view municipal water management. It is important to note that water demand management implies only an increased use of instruments that are designed to provide effective water use incentive systems. The term does not imply direct public intrusion into what is essentially private individual or company decision-making

Water demand management views the use of water as a variable “demand” in the economic sense - that is, as a variable that is responsive to price and its determinants, as well as other socio-economic variables. In other words, a demand management approach views water as no different from most other commodities in a modern economy. This contrasts sharply with the view commonly held that water is a “requirement” that must be met, frequently regardless of cost. It must be stressed that water demand management is an adjunct to, not a replacement for, current approaches. Basically, it involves a slight re-orientation of thinking, not a wholesale “revolution” in procedure.

4.9.1 A Definition of Water Demand Management

Formally defined, water demand management refers to “any socially beneficial measure that reduces or reschedules average or peak water withdrawals from surface or ground water sources while maintaining or mitigating the extent to which return flows are degraded.” (Tate, 1990) This definition contains four basic concepts that merit a brief discussion here because they run implicitly through most of the following chapters. These four concepts are as follows;

socially beneficial refers to the requirement that measures undertaken to manage water demands should show an excess of (social) benefits over (social) costs. Using this concept, demand management applies to any type of project to improve the efficiency of water use, regardless of whether or not actual shortages of water exist in any particular area, providing that the benefits outweigh the costs.

reducing or rescheduling average or peak water demands refers to the different conditions under which demand management actions might occur. Average and peak demands are the two most common determinants of water system size, and accordingly the need for investment. Actions that reduce either or both of these flow characteristics will have long-run impacts on system investment.

surface or ground water emphasizes that both water sources are important in the management of water demands. This may appear somewhat self-evident, but, in applied practical terms, surface and ground water are often considered in isolation from each other. This part of the definition stresses that both sources are equally important, and further, need to be approached in an integrated manner during the water management process.

maintaining or mitigating the quality of return flows stresses that demand measurements should not lead to water quality deterioration. In other words, demand management measures should be at least benign with respect water quality, or lead to quality improvements.

4.9.2 Principal Components of Water Demand Management

At the heart of demand management lies a body of *economic techniques and principles*, which forms the first element of the framework. Economic markets form one of the major institutional underpinnings of modern Western economies. As outlined earlier in the report, while much of the basic economic theory is based on a model of perfect competition, these economies, in fact, are “mixed” to greater or

lesser degrees, in the sense that they combine public, non-market arrangements with private markets to achieve the allocation of resources, goods, and services. Thus, while modern economies are often far from the model economies of textbooks, they do employ economic markets as central means of allocating resources in modern Western societies, serving as the major allocative mechanism for most resources, goods and services traded or used. Water demand management uses these market or quasi-market conditions, along with other “non-structural” means in an effort to modify current water use patterns to achieve improved water resource sustainability.

The second element of the conceptual framework is also socio-economic in nature – namely a set of *socio-political characteristics and techniques* that are central to the way in which societies operate. These include the basic framework of law and institutions that form much of the fabric of entire countries. Legal and institutional arrangements form the basis of the “rules” that allow economies to operate, and to allocate property rights to resources. It follows that laws and institutions play a significant role in governing the ways in which water is used, and how it is controlled and managed. To change basic methods and patterns of water demand requires varying degrees of alteration to basic laws and institutions. For example, altering Western North American water allocation patterns would entail dealing with a well established and fairly entrenched system of seniority rights to the use of water. Modifying this system of water rights has proven a challenging and often insurmountable task. More pertinent to Ontario would be changes in regulatory approaches to allow effluents charges or economics-based charges for taking water. A second set of socio-political instruments concerns public information and education. A wide range of research has shown concerted efforts to inform and educate the public as to wise water use can result in substantial decreases in the demand for water (Nieswiadomy, 1992). Finally, a wide range of public policy is included in the socio-political dimension of water demand management. These policies may be reflected explicitly, for example, in government fiscal policies, or implied in policy statements such as those enunciated by many governments supporting sustainable development or other broad goals.

Although demand management emphasizes economic and socio-political actions, ultimately these must be reflected in modifications to physical works and equipment. A third major element of a demand management framework focuses on a range of *technical, structural, and operational techniques*. Structural measures include measures such as: system leakage and infiltration prevention, metering, retrofitting, recycling, and other measures. All of these measures are related to improving the physical efficiency of water use. Operational actions, those relating to the procedures used to operate water systems, include issues such as measurement of water demands and planning.

4.9.3 The Importance of Water Demand Management

Water demand management is important for several reasons, among which are the following:

- The approach highlights the finding that water use is alterable through pricing and other non-structural means. This is a valuable insight in the long run because it implies that improved efficiency in water use and patterns can influence capital investment levels, which will lead, in turn, to lower requirements for water infrastructure spending.
- The basis for achieving economic efficiency in municipal infrastructure is the measurement and comparison of benefits and costs of decisions made with respect to water servicing provision. Demand management, by definition, include a focus of “socially beneficial” actions and decisions, thereby requiring implicitly both the measurement of demands and the conduct of benefit cost analyses.
- In planning water servicing expansions or major modifications, system planners should be required to take all alternatives into account to determine whether or not major works can be sized differently, altered with respect to timing, or otherwise changed to lower cost. Demand side management approaches encourage such a consideration of alternatives.

- Demand management identifies new alternatives that may help in planning future system modifications.

4.10 Summary

For the purposes of this report, water demand referred to both the demand for water supplies and wastewater treatment. Records from the 1989 to 1999 period demonstrate that municipal demands have grown in Ontario, but still for a small proportion of total water demands. Municipal water demands are not evenly spread through time, but show pronounced peaks on a daily and seasonal basis. These peaking characteristics could form important considerations in setting water rates, particularly pricing based on predictable seasonal patterns of use.

Water is a “normal” good in the economic sense, in that as price rises, demand falls. A large number of empirical studies have confirmed this characteristic. This is a key finding for public policy. Demand responds to price changes in various ways, depending on the type of demand, the time of year, the short- versus the long-run, and other factors. The relationship is measured by the concept of price elasticity of demand, which generally reflects the availability of substitutes to the use under consideration. In general, the fewer the substitutes, the lower the price elasticity. Many empirical studies have permitted price elasticities to be defined within relatively narrow ranges. Generally, most municipal water demands are price inelastic, with the most inelastic values (i.e., the lowest) occurring for indoor demands (a reasonable range for practical purposes is suggested at -0.2 to -0.4). Outdoor elasticities are relatively higher (-0.4 to -0.6), with industrial elasticities often higher still.

Econometric studies in the municipal water demand have tended to concentrate in the residential demand area, with much less attention paid to commercial, (publicly-supplied) industrial, and public demands. Prices of other inputs also have an impact on water demand, especially prices of waste treatment and energy.

Income levels have a positive impact on residential water demands, in that as average income levels rise, so also do water demands. Income elasticity values are in the range of (+) 0.4. This factor tends to be of less interest in water demand studies because water managers have little chance to influence income levels.

Metering water demands can have a substantial impact on water demands. Complete metering in a municipality, combined with appropriate pricing, has been shown to lower demand levels by up to 50% over flat rate, unmetered uses.

Water demand management presents a philosophy of management that focuses on the demands made on water resource systems. Demand management can have significant impacts on system costs by fostering lower levels of demand, and thereby lower system capital and O&M costs.

Appendix 4.1: Price and Income Elasticities

Table A4.1 Price and Income Elasticities in North America ¹							
Year	Author	Study Location	Type of Use	Type of Data	Type of Model	Own-Price Elasticity	Income Elasticity
1951	Larson and Hudson	15 communities in Illinois	M & I	CS			0.7
1957	Seidel and Baumann	111 areas in the U.S.	M & I	CS	L	-0.12 to -1.00	
1958	Fourt	34 U.S. cities	M & I	CS	LL	-0.39	0.28
1963	Bain	Northern California	M & I	CS	LL	-1.1	0.58
1963	Gottlieb	Kansas	M & I	CS	LL	-0.67 to -1.23	0.28
1963	Headley	San Fransisco Bay Area	M & I	TS	L		0.4
1964	Gardner-Schiek	Northern Utah	M & I	CS	LL	-0.77	
		Northern Utah	M & I	CS	L	-0.67	
1965	Flack		M & I	CS	I	-0.12 to -0.61	
1966	Ware and North	Georgia households	SF Residential	CS	LL	-0.61	0.38
		Georgia households	SF Residential	CS	L	-0.67	0.83
1967	Conley	So. California	Agg. Residential	CS		-1.02	
1967	Howe and Linaweaver	21 areas across the U.S.	SF Residential, annual	CS	L	-0.4	0.47
		21 areas across the U.S.	SF Residential, winter	CS	L	-0.23	0.32
		11 areas in the east 11 areas in the west	SF Residential, summer	CS	L	-1.57	1.45
		10 areas in the east 10 areas in the west	SF Residential, summer	CS	L	-0.73	0.69
1969	Turnovsky	19 Massachusetts towns	Agg. Residential	CS	L	-0.25, -0.28	
1970	Male, Willis, Babin, and Shilito	Northeast US	Agg. Residential	CS	L	-0.2	
		Northeast US	Agg. Residential	CS	LL	-0.37	
	Hittman Assoc.	United States	Agg. Residential	CS	LL	-0.44	
1971	Young	Tucson, Arizona	M & I	TS	L	-0.65	
		Tucson, Arizona	M & I	TS	LL	-0.69	
1972	Wong	City of Chicago, 1951-61	M & I	TS	LL	-0.02	0.2
		Chicago suburbs, 1951-61	M & I	TS	LL	-0.28	0.26
		Chicago suburbs, Over 25,000, 1961	M & I	CS	LL	-0.53	1.03
		Chicago suburbs, 10-25,000, 1961	M & I	CS	LL	-0.82	0.84

		Chicago suburbs,	M & I	CS	LL	-0.46	0.48
		Chicago suburbs, Under 5,000, 1961	M & I	CS	LL	-0.27	0.58
1972	Grima	Toronto, Ontario	SF Residential, Annual use	CS	LL	-0.93	0.56
		Toronto, Ontario	SF Residential, Winter use	CS	LL	-0.75	0.41
		Toronto, Ontario	SF Residential, Summer use	CS	LL	-1.07	0.51
1973	Morgan	92 residences in Santa Barbara County	SF Residential, Winter use	CS+TS, monthly	L	-0.49	0.53
1974	Primeaux and Hollman	402 households in 14 Mississippi cities	SF Residential	CS	LL	-0.45	0.24
		402 households in 14 Mississippi cities	SF Residential	CS	L	-0.37	0.26
1974	Sewell and Roueche	17 areas in Victoria, B.C., 1954-1970	M & I	CS+TS	L	-0.46	0.27
		17 areas in Victoria, B.C., 1954-1970	M & I	CS+TS	LL	-0.39	0.19
1975	Andrews and Gibbs	Miami, Florida	M & I	CS	SL	-0.62	0.8
1975	Hogarty and Mackay		SF Residential, short-run	CS+TS	L	-0.86	
			SF Residential, long-run	CS+TS	L	-0.56	
1975	Pope, Stepl, Lytle	South Carolina households, 1965- 71	SF Residential, irrigators	CS+TS	L	-0.31 to - 0.67	
		South Carolina households, 1965- 71	SF Residential, Non-irrigators	CS+TS	L	-0.06 to - 0.36	
1976	Grunewald, Haan, Debertin, and Carey	Kentucky	Agg. Residential	CS	LL	-0.92	
1976	Morgan and Smolen	33 areas in Southern California	M & I, annual	CS	L	-0.44	0.33
		33 areas in Southern California	M & I, winter	CS	L	-0.45	
		33 areas in Southern California	M & I, summer	CS	L	-0.43	
1977	Clark and Goddard	22 areas	M & I	CS	L	-0.63	
		22 areas	M & I	CS	LL	-0.6	
1977	Gallagher and Robinson		SF Residential, winter	CS+TS	LL	-0.24	
1977	Gardner	Minnesota areas	Agg Residential	CS	L	-0.24	
		Minnesota areas	Agg Residential	CS	LL	-0.15	
1977	Danielson	Raleigh, North Carolina	M & I, annual	TS		-0.27	
		Raleigh, North Carolina	M & I, winter	TS		-0.305	

		Raleigh, North Carolina	M & I, summer	TS		-1.38	
1978	Camp	288 households in 10 Mississippi cities	SF Residential	CS	L	-0.24	
1978	Carver	Fairfax, Co, 1974-75	Agg Residential, Summer	TS, monthly	L	-0.13 to -0.17	
		Fairfax, Co, 1974-75	Agg Residential, winter	TS, monthly	L	-0.02 to -0.04	
1978	Gibbs	Households in Miami, Florida	SF Residential	CS+TS	L	-0.51 to -0.61	0.51 to 0.8
1979	Cassuto and Ryan	Oakland, 1970-75	Agg Residential	TS+CS	L	-0.14 to -0.3	

1979	Danielson	261 households in Raleigh, NC, 1969-74	SF Residential, annual	CS+TS	LL	-0.27	0.33
		261 households in Raleigh, NC, 1969-74	SF Residential, winter	CS+TS	LL	-0.3	0.35
		261 households in Raleigh, NC, 1969-74	SF Residential, summer	CS+TS	LL	-1.38	0.18
1979	Foster and Beattie	217 U.S. cities	M & I	CS	SL	-0.53	0.18
		New England	M & I	CS	SL	-0.43	
		Midwest	M & I	CS	SL	-0.3	
		South	M & I	CS	SL	-0.38	
		Plains	M & I	CS	SL	-0.58	
		Southwest	M & I	CS	SL	-0.36	
		Pacific Northwest	M & I	CS	SL	-0.69	
1979	Male et al.	Eastern United States	M & I	CS	L	-0.2	0.25
		Eastern United States	M & I	CS	LL	-0.68	0.46
		Eastern United States	M & I	CS	SL	-0.35	0.55
1980	Agthe and Billings	Tucson, Arizona 1974-77	Agg Residential	TS	L	-0.18, -0.36	
		Tucson, Arizona 1974-77	Agg Residential	TS	LL	-0.18, -0.26	
1980	Ben-Zvi	Red River Basin	Agg Residential, Annual	CS	LL	-0.73	
		Red River Basin	Agg Residential, Winter	CS	LL	-0.79	
		Red River Basin	Agg Residential, Summer	CS	LL	-0.82	
1980	Billings and Agthe	Tucson, Arizona 1974-77	Agg Residential	TS	LL	-0.27	
		Tucson, Arizona 1974-77	Agg Residential	TS	L	-0.49	
1980	Carver and Bolland	13 areas	M & I, winter, long-run	TS+CS	L	-0.7	
		13 areas	M & I, winter, short-run	TS+CS	L	-0.5	
		13 areas	M & I, summer, long-run	TS+CS	L	-0.11	
		13 areas	M & I, summer, Short-run	TS+CS	L	-0.1	
1980	Morris and Jones	Households	SF Residential, Annual	CS	LL	-0.39	
		Households	SF Residential, Winter	CS	LL	-0.09	
		Households	SF Residential, Summer	CS	LL	-0.73	
1981	Foster and Beattie	United States	M & I	CS	SL	-0.47	0.46
1981	Hansen and Narayanan	Salt Lake City, 1961-77	M & I	TS	LL	-0.47	
1982	Billings	Tucson, Arizona 1974-77	M & I	TS	L	-0.66	

		Tucson, Arizona 1974-77	M & I	TS	LL	-0.56	2.14
1982	Howe	United States United States, east	SF Residential, Summer	CS CS		-0.06 -0.57	
		United States, west	SF Residential, Summer	CS		-0.43	
1982	Morgan	473 households in Oxnard, California	SF Residential	CS+TS monthly	L		0.46
1984	Jones and Morris	326 households	SF Residential	CS	LL	-0.21	
1987	Moncur	1281 households in Honolulu	SF Residential	CS+TS monthly	L		0.38 to 0.8
1989	Billings and Day	11 districts in Tucson, Arizona, 1974-1980	Agg Residential	CS+TS			0.36
1989	Weber	12 sub-districts of EBMUD, Oakland	Agg Residential, winter	TS+CS	L	-0.2	
		12 sub-districts of EBMUD, Oakland	Agg Residential, winter	TS	L	-0.08	
1990	Boland, McPhail, and Opitz	Households in Southern California	SF Residential, Winter	CS+TS	LL	-0.01 to – 0.02	
1991	Dziegielewski and Opitz	Households in Southern California	SF Residential, Summer	CS+TS	LL	-0.13 to – 0.18	
		Metropolitan Water District of Southern California	SF Residential, Summer	TS+CS	LL	-0.24	
		Metropolitan Water District of Southern California	SF Residential, Winter	TS+CS	LL	-0.39	
		Metropolitan Water District of Southern California	MF Residential, Summer	TS+CS	LL	-0.13	
		Metropolitan Water District of Southern California	MF Residential, Winter	TS+CS	LL	-0.15	
SOURCE: Hannemann, as contained in Baumann <i>et al.</i> 1998							
KEY: SF = single family; MF= multifamily; Agg = aggregate; CS = cross-section; TS = time series; L = linear demand; LL = log-log demand; SL = semilog demand.							

CHAPTER 5 : WATER SUPPLY AND ASSESSMENTS OF WATER UTILITY COSTS

In a manner directly analogous to the analysis of demand, the other half of the water servicing issue relates to the economics of supplying water services. Here the emphasis is on the costs involved in carrying out this task. This chapter describes the principal dimensions of the supply problem of water utilities, focusing on two main issues: (a) the methodology used to analyze costs from a practical, engineering viewpoint at the master planning level; and (b) econometric analyses that have been carried out on the supply/costing. The latter are normally based upon ordinary least squares (OLS) regression or variations upon it.

Supply factors are central to the determination of the costs of meeting current and future water demands. Overall system size is ultimately driven by the interaction of supply and demand factors, but, as we have already shown both historically and currently, these two sets of factors have been somewhat divorced from one another. Their integration is of considerable importance in meeting the financial challenges of the future, particularly in the setting of water rates. The integration issue is addressed in chapter six.

5.1 Engineering Practice in Water Servicing: An Introduction⁴⁴

This chapter begins with an overview of current engineering practices in Ontario with regard to planning and costing of water and wastewater infrastructure. This discussion is intended to provide insights into the basic standards that govern the supply of water services, in terms of both water supply and waste treatment. It also outlines the major guidelines typically used to size infrastructure. In terms of cost estimation, this section describes common engineering rules of thumb, and identifies typical factors which influence the cost of construction, including location, schedule, economic conditions, and time of year. Descriptions of standard quantitative relationships are included, as are lists of engineering text references commonly used to estimate costs for specific construction projects. This discussion provides the foundation for dealing with the economic problems of supplying municipal water services.

5.1.1 Major Water Supply System Components

In chapter one, a brief overview was provided of a generic municipal water servicing system. The following brief discussion amplifies that overview, and places it within an engineering context.

Municipal water systems consist basically of three principal components:

a. **Raw Water Intake Facilities**

Raw water intake facilities are used to withdraw water from the supply source, which can be either surface-water sources (lake, river, etc.) or a groundwater source drawing from underground aquifers or springs. Facilities for surface water sources normally include an intake pipe installed into the lake or river, and a crib-like structure to protect the mouth of the pipe. The supply facilities may

⁴⁴ Michael Hribljan and Leanne Jones, of Acres Associated Environmental Limited, prepared this section. Some of this material was presented in Chapter One in a slightly different form, and is reviewed here to provide the correct engineering context for this chapter.

also include screening and pumping facilities to convey the water supply through other components associated with the system. Groundwater supply facilities include wells or a network of perforated pipes or chambers installed into the source aquifer used for water supply. The overall aim is to obtain the best quality water possible for the municipality, with the complexity of the supply facilities being dictated by the level of effort required to carry out this task. AWWA Water Treatment Plant Design reports that the intake system may represent 20% of the total water treatment plant investment.

b. Water Treatment Facilities

Provincial regulations require that all municipal water supplies receive a minimum level of treatment before delivery to the consumer to remove contaminants from the source water supply to ensure potability. In Ontario, dual water systems that separate potable from non-potable supplies are rarely used. Thus, all publicly supplied water is required to meet potable standards. Treatment of raw water depends on its quality and can range from simple disinfection for inactivating microbial pathogens in the case of higher quality groundwater supplies to advanced processes for particulate matter removal, taste and odour control and advanced disinfection for poorer quality supplies from surface sources.

Most often, the treatment of raw water includes both physical removal of particulate matter and removal/inactivation of micro-organisms. The latter is accomplished through chemically assisted coagulation/flocculation, filtration and disinfection. "Full (or Conventional) Treatment" includes chemically assisted coagulation/flocculation, sedimentation, filtration plus disinfection. Where source water quality allows, the sedimentation step may be avoided in a process defined as "Direct Filtration". Either facility design will include a "clearwell" which is a reservoir provided for the storage of filtered water and provision of contact time for disinfection.

For water supplies requiring higher levels of treatment, other processes may supplement or replace the above-noted facilities including, but not limited to ozonation, ultra-violet irradiation, microfiltration, reverse osmosis, and aeration. These more advanced treatment methods substantially increase supply costs.

c. Water Pumping and Distribution

The purpose of the pumping and distribution system is to deliver an adequate quantity of treated water supply at sufficient pressure to the consumer. The system is comprised of the pumping facilities, storage (both elevated or floating and in-ground), transmission pipelines and distribution water mains. Depending on the system, additional disinfection may be provided within the distribution system to maintain water quality at potable levels.

5.1.2 Major Waste Treatment System Components

Municipal wastewater systems typically consist also of three major components:

a. Wastewater Collection and Pumping

Wastewater collection systems consist of a network of underground pipes, conduits, tunnels, equipment and appurtenances for the collection, transportation, and pumping of wastewater. Normally, wastewater flows through the system via gravity or a combination of gravity and pumping. Pumping stations can be equipped with screening, odour control, and standby power. There are three main types of municipal sewers: sanitary sewers, storm sewers and combined sewers. Sanitary sewers receive wastewater from residential, commercial, institutional, or industrial sources, as well as small quantities of groundwater infiltration or inflow. Storm sewers convey

stormwater runoff and other drainage only, while combined sewers convey both sanitary wastes and stormwater.

b. Wastewater Treatment Facilities

Regulations govern the effluent quality that must be achieved by waste treatment facilities prior to discharge into receiving waters. Treatment needs are dependent on the sensitivity of the receiving body and its assimilative capacity. Typically, in Ontario, a minimum of secondary treatment or equivalent is required. The treatment process includes physical, chemical and biological processes to remove pollutants from the wastewater. Common unit processes include raw wastewater pumping, screening and grit removal, primary sedimentation, biological treatment, secondary clarification, filtration, disinfection, and digestion.

c. Wastewater Disposal

The method of effluent discharge is influenced by the characteristics of the receiving body and the effluent quality. Discharge methods include direct discharge to receiving waters through engineered diffusers, percolation, wetlands, land application, deep-well injection, and groundwater recharge.

Following treatment, biosolids generated at the facility must be disposed removed. Common disposal/reuse methods include beneficial agricultural land application and landfill.

5.2 Cost Estimating Methodology

5.2.1 Types of Estimates

The accuracy of cost estimates is dependent on the amount of data available to characterize the site and the project requirements. Reliable construction cost data is essential for planning, design, and construction of any waterworks project. As the project develops, more details become available, and it is possible to produce a more accurate cost estimate. The American National Standards Institute (ANSI) and American Association of Cost Engineers (AACE) recognize the following three levels of cost estimates:

- Study or Order of Magnitude Estimate
- Preliminary or Budget Estimate
- Pre-Tender or Definitive Estimate

A level of uncertainty is associated with each type of cost estimate. Uncertainty or unknowns within a cost estimate are usually addressed through the inclusion of allowances and contingencies. Allowances are incremental monies included in the estimate to cover known, but undefined requirements, such as provisional items. Contingencies provide for unforeseeable elements that are not within the scope of the project.

The following overview provides a description of each type of cost estimate, including the situation in which it is normally used.

a. Study or Order of Magnitude Estimate

A study or order of magnitude estimate is typically used in the planning stage of a project to compare and evaluate process alternatives, or establish order of magnitude budget costs. This type of estimate is suitable for costing at a master plan level, and is applicable to projects in Phases 1 or 2 of the Municipal Class Environmental Assessment planning process. As a result, the estimate must be accurate enough to allow sound decisions to be made regarding alternative selections. The AACE anticipates that a study or order of magnitude estimate should be accurate to within plus 50% or minus 30%. At this point in the project, the level of detail available is generally very limited,

and accordingly standard cost factors or cost curves are normally used. Cost estimating software can also be used to produce order of magnitude estimates.

Cost factors and curves vary by region and time period, requiring that estimates be adjusted to account for the resulting differences. The most common method used to adjust cost estimates is the cost index. A cost index is a calculated numerical value that is a function of an established quantity of material and labour. The most commonly used single indices in the construction industry are the Engineering News Record (ENR) construction cost index (CCI), the ENR building cost index (BCI) and the Southam construction cost index. Although single indices provide an uncomplicated means of adjusting estimates, they are often inadequate for application to the construction of water and wastewater infrastructure. To overcome the shortcomings, total construction costs are divided into eight major cost components, and the total construction cost is updated based on eight principal cost components by using appropriate indices, rather than a single index.

At the planning stage, the design is based on assumed engineering parameters. In order to accommodate additional costs that may arise due to a change in these parameters, an information contingency allowance of approximately 15% is typically used.

b. Preliminary or Budget Estimate

A preliminary or budget estimate is prepared with the use of flow sheets, layouts, and equipment details. The level of information available for preliminary estimates should fall between that available at the study stage and that which will be available at the pre-tender stage. This type of estimate is intended for a client's budget, and as such, the AACE anticipates that the estimate should be accurate to within plus 30% or minus 15%. Normally, a preliminary or budget estimate is prepared based on approximate quantities, site specific data and conditions, and manufacturer's cost data (if applicable).

At the preliminary design stage, detailed engineering drawings are not yet available, and the design is still based on a number of assumed engineering parameters. In order to accommodate additional costs that may arise due to a change in these parameters, an information contingency allowance of approximately 10% is typically used.

c. Pre-Tender or Definitive Estimate

A pre-tender or definitive estimate is developed during the detailed design phase of the project, and is used to compare and evaluate contractor bids. The AACE anticipates that this type of estimate should be accurate to within plus 15% and minus 5%. Pre-tender or definitive estimates are influenced by a number of factors, including system capacity, site conditions, climate, permit costs, and local and nationwide economic conditions.

At this point in the project, all project specific requirements, such as type and quantity of materials, workmanship, quality of finishes, location, schedule, etc. should be known. Any unknown details will reduce the accuracy of the estimate. Engineering data available at this stage should include at a minimum, complete plans and elevations, piping and instrumentation diagrams, single line electrical diagrams, equipment data sheets and quotations, structural sketches, soil data and sketches of major foundations, building sketches, and a complete set of specifications and quantities.

In order to account for unforeseen construction difficulties, a safety factor should be included in the estimate as follows:

- 2% where working drawings are in the final stage
- 7% where working drawings are in the preliminary stage

- 10-20% for alterations or repairs

A contingency allowance for unforeseen design and construction changes for which the contractor will be paid extra should also be included in the estimate. This contingency allowance is typically shown on the bid form, and all changes must be authorized by the engineer prior to payment. This contingency allowance is dependent on the size and complexity of the project. Table 5.1 identifies suggested contingency amounts based on project cost.

Table 5.1 Recommended Contingency Allowance	
Percent Contingency on Base Cost	Project Base Cost
10%	Up to \$100,000
8%	\$100,000 to \$500,000
6%	\$500,000 to \$1,500,000
4%	\$1,500,000 to \$3,000,000
2%	Over \$3,000,000

Two standard approaches to preparing a pre-tender or definitive cost estimate are commonly used: the bid price database method; and the labour and materials method. The selection of the method to be used should be based on the requirements of the project and the information available. For some projects, a combination of the two methods may be most appropriate.

The bid price database method requires a detailed list of quantities, and a database of historic bid prices for similar work. Quantities can be taken from detailed drawings and should be adopted in a manner consistent with contract measurement clauses. Unit prices can be obtained from a database of historic bid prices for similar work. It is preferable that the previous work be recently completed and of similar scale to the proposed undertaking. It is recommended that two or three previous contracts be utilized to determine a unit price for each work item. Unit prices on contracts generally include overhead and profits. Therefore these items do not need to be specifically addressed in the cost estimate. To account for inflation, an inflation factor obtained from a cost index is usually used. If the database of historic bid prices does not include an item for similar work, a unit price can be calculated based on material and labour costs. Alternatively, contractors or suppliers may be contacted for a quote.

The labour and materials method requires a detailed list of quantities, material costs, productivity rates, and labour and equipment costs. Detailed drawings will provide these quantities. Product suppliers can provide the most accurate estimates of material costs. It is important to recognize that price varies with quantity, and some contractors will be given preferential pricing. Productivity rates, labour and equipment costs associated with specific work items are difficult to accurately assess. Information is available through a number of sources including equipment performance charts, Means Cost Data, and Richardson's Estimating Standards. Values for equipment and labour can also be obtained from reference material, such as labour agreements, equipment rental guides and account sheets from other contracts.

Additional costs not associated directly with a work item must also be included in a pre-tender or definitive cost estimate. These costs may include contractor's superintendent, site office, bonding and insurance, and mobilization and demobilization. Normally, such costs are included through the addition of standard percentages or calculated directly.

5.2.2 Indirect Costs

In addition to constructions costs, other project costs should be considered when preparing a cost estimate include engineering fees, administration and legal costs, land acquisition costs, and provincial and federal taxes. The allowance for these items is dependent on the size and

complexity of the project. In general, lower allowances apply to larger projects without complicated mechanical systems, while higher allowances apply to smaller, more complex projects or projects with a high degree of mechanical complexity.

a. Engineering Fees

Engineering fees account for any engineering services that are to be provided and may include special investigations, surveys, foundation explorations, location of interfering utilities, inventory of existing facilities, pre-design, design, site inspection, preparation of construction drawings and specifications, construction management and inspection, materials testing, final inspection, start-up and commissioning, and preparation of record drawings. Engineering fees typically range between 5% and 25% of project construction costs depending on the magnitude and complexity of the project. 5% is typical for less complex projects, such as buried pipelines, while 25% is typical for low capacity, complex treatment works.

b. Administration and Legal Fees

Administration costs include in-house engineering and administrative costs, as well as fees for approvals and permits. Experience has shown that administration and legal costs can vary from 12-15% for externally designed projects, to 20% for in-house designed projects. As a general rule, an average allowance of 15% of the total construction cost is recommended to cover administrative and legal costs.

c. Land Acquisition Costs

Land acquisition costs are specific to each project and should be evaluated based on individual project requirements. A number of factors can directly impact land costs including site location, size, zoning, and economic conditions.

d. Provincial and Federal Taxes

Provincial tax is charged only on the materials required for the project, not for the services received. Tax on construction materials is already accounted for in the cost data used to develop unit cost relationships, and therefore does not need to be included as an additional item. All project components, including construction cost, engineering fees, administrative costs, contingencies and land costs are subject to the Federal Goods and Services Tax (GST). Typically a net rate of 3% is applied for municipal projects as municipalities receive a 4% GST rebate. However, the rate and applicability of the tax is dependent on the client, and the nature and location of the services to be provided.

5.3 Costing - Part A: Design Standards

5.3.1 Sizing Criteria

The Ontario Ministry of the Environment (MOE) Guidelines for the Design of Water Treatment Plants and Sewage Treatment Plants (1984) provides recommendations for the sizing of water and wastewater infrastructure in Ontario. Due to the age of the guidelines and the limited technology, Water Environment Federation (WEF) Manual of Practice 8 (MOP 8) is the recognized industry standard for wastewater design in Ontario. The "10 State Standards" are commonly used as a reference for water system design.

Specific criteria related to sizing of water and wastewater infrastructure is covered in the following section. The criteria are conservative in nature, and therefore can be useful at the master planning level, if applicable.

5.3.2 Water Supply, Treatment, Pumping and Distribution

In Ontario, water supply systems are generally designed to satisfy the greater of either maximum daily demand plus fire flow requirements or the peak rate, projected over a 20-year period. 10-year design periods can be used for larger capacity systems, where cost may be the overriding factor, and 50-year or ultimate design periods are generally used for components where the cost of the work is not as substantially dependent on the size used (e.g., intakes and large transmission mains). The maximum daily demand is defined as the average usage rate on the maximum day throughout the year, while the peak rate is the average water usage over the maximum hour.

Water demand varies from municipality to municipality depending on many factors including demographics, water rate structure, by-laws for outdoor usage, and system leakage rates, as discussed in chapter four. It is recommended that, wherever possible, historical records be used as a basis for projections; however, in their absence, the MOE recommends average daily domestic per capita demands of 270-450 Lpcd and peaking factors based on the size of the municipality. Industrial, commercial and institutional demand projections should also be based on historical records; however, recommended unit water demands are also available for design purposes. For example, common allowances for light industry are 35 m³/gross hectare-day with peaking factors of 2 to 4 times the average usage rates.

The reader will note the contrast between the material presented here, and the contents of chapter four. We have included this description because it is typical of engineering practice currently. One of the implicit challenges for water rate making in the future will be to reconcile these different practices and criteria.

Fire flow requirements are generally estimated based on the latest version of "Water Supply for Public Fire Protection – A Guide to Recommended Practice", prepared by Fire Underwriters Survey.

Water treatment plants are sized to provide the maximum day water demand with the balance of peak demands provided from storage facilities within the system. The treatment process capability of the plant must; however be greater to account for water used in processes such as filter washing, ozone generator cooling water, service water, chlorine ejectors, etc. Historical records and manufacturer's data are the most accurate way to calculate the allowance for in-plant usages; however a 6% factor is used in most typical designs. Allowances in design may also be necessary to account for seasonal variations in raw water quality. Low water temperatures can result in longer reaction times and lower treatment efficiency in sedimentation tanks and filters which will make it more difficult for the plant to meet its design conditions during those periods.

Table 5.2 summarizes the design criteria used for each major treatment component in the water system assuming a 6% allowance for in-plant usage. For redundancy, facilities are designed so that rated or firm capacity can be met with the largest unit of out service. It should be noted that chemical feed systems typically have a shorter design life due to their relatively shorter life-span and lower capital costs. It can also be difficult to operate these systems accurately at the lower end of their design flow ranges.

Table 5.2: Sizing Criteria for Water Works Infrastructure		
Component	Sizing Criteria	Design Year
Intake	106% of Maximum Day Demand	50 year
Raw Water Pumping & Screening	106% of Maximum Day Demand with the largest pumping unit of out service	20 year
Mixing/Settling	106% of Maximum Day Demand	20 year
Filtration	108% of Maximum Day Demand (assumes that filters operate 23.5 hrs/day to allow for downtime during backwash)	20 year
Clearwell	10-20% plant capacity, provided that sufficient volume is provided at all times for disinfection	20 year
High Lift Pumping	With floating storage: Maximum Daily Demand with the largest pumping unit of out service Without floating storage: Greater of Maximum Daily Demand + Fire Flow or Peak Hour with the largest pumping unit of out service	20 year
Chemical Feed Facilities	Maximum chemical dosage required at 106% of Maximum Day Demand	10 year
Distribution System Storage	Total Storage = Balancing storage (25% of Maximum Daily Demand) + Fire Storage + Emergency Storage (25% of Balancing + Fire)	20 year
Distribution Booster Pumping	With floating storage: Maximum Daily Demand with the largest pumping unit of out service Without floating storage: Greater of Maximum Daily Demand + Fire Flow or Peak Hour with the largest pumping unit of out service	20 year

5.3.3 Wastewater Collection, Pumping, Treatment, and Disposal

The rated capacity of wastewater treatment plants is generally based on the average daily flow rate projected over a 20-year design period. The average daily flow rate is defined as the average flow rate occurring over a 24-hour period based on total annual flow data. The peak flow rate is defined as the peak sustained hourly flow rate occurring over a 24-hour period based on annual operating data. Peak hourly flow rates are typically used for the design of collection and interceptor sewers, wastewater pumping stations, grit chambers, sedimentation tanks, chlorine contact tanks, and conduits or channels in the treatment plant.

The MOE recommends that design flow rates and loadings be determined based on historical data. However, if historical data is not available, a wastewater generation rate of no less than 225 Lpcd may be used for average domestic wastewater flows, exclusive of extraneous flows. Extraneous flows, including infiltration and inflow, vary from municipality to municipality; however, the MOE guidelines recommend that no less than 0.1 (L/mm-d)/m (litres per millimeter of sewer diameter per day per linear metre of total sewer system length including sewer connections) be used to estimate peak extraneous flows. The peak flow rate can be determined by multiplying the average domestic flow rate by the Harmon peaking factor, then adding the peak extraneous flow. Industrial, commercial and institutional wastewater flow rates should be based on historical records, but may also be estimated based on empirical values.

Typically, all components of wastewater treatment plants should be hydraulically capable of handling the anticipated peak flow rates without overtopping (Table 5.3). The Ontario Ministry of the Environment (MOE) Guidelines for the Design of Water Treatment Plants and Sewage Treatment Plants (1984) recommends sizing of specific unit processes based on the following hydraulic, organic and inorganic loading rates. For further details, refer to the abovementioned document.

Table 5.3 Sizing Criteria for Wastewater Infrastructure	
Component	Sizing Criteria
Wastewater pumping station	Peak hourly flow rate
Screening	Peak hourly flow rate
Grit removal	Peak flow rate, peak grit loading rate
Primary sedimentation	Peak flow rate, peak suspended solids loading rate
Aeration (w/o nitrification)	Average diurnal BOD ₅ loading (usually sufficient)
Aeration (w/ nitrification)	Average diurnal BOD ₅ loading (usually sufficient for predominantly domestic wastes)
Secondary sedimentation	Peak flow rate and/ or peak solids loading rate (varies with treatment)
Sludge return	Varies with treatment
Disinfection	Peak flow rate, unless downstream receiving stream dictates otherwise
Chemical feed	Peak flow rate
Effluent filtration	Peak flow rate, peak solids loading rate
Outfall sewer	Peak flow rate
Sludge Treatment	Average loading rates (hydraulic, total solids, volatile solids) unless sustained peaks are significant to treatment process

The overall objective of the design is to provide a wastewater treatment system capable of handling a wide range of wastewater flows and loadings, while complying with the overall performance requirements. In order to achieve this, the influence of varying flows and loads must be considered.

5.3.4 Stormwater Collection & Disposal

Sizing of stormwater collection and disposal systems is dependent on the size of the watershed, the duration and frequency of the design storm, the volume of runoff, and other potential sources of flow, such as groundwater infiltration. Due to the difficulty associated with estimating stormwater volumes, computer models, such as Storm Water Management Model (SWMM), are often used. Models are typically calibrated with rainfall data collected from one storm, and the calculated results are compared with the observed field measurements. The estimated input parameters are adjusted to obtain the best fit between measured and predicted volumes, producing the volume of flow that is conveyed through the system. The sizing of collection pipes and outfalls can then be determined.

5.3.5 Technology Selection

Having identified the capacity of the systems required, the next step in planning is to identify the type of treatment processes required for the system. Two components factor into this decision: ambient water quality; and the treatment required to meet or exceed regulatory standards or objectives, both now and in the future. For instance, most water treatment facilities over the last 10 years have designed disinfection facilities to a higher standard than provincially regulated, as it was rightfully believed that the standards would become more stringent as the provincial regulatory authority acknowledged research being carried out in other jurisdictions.

5.3.6 Ambient Conditions

a. Water Supply, Treatment, Pumping & Distribution

Table 5.4 summarizes the effect of ambient water quality on the level of treatment to be provided for surface water systems, based on a reference provided in *Water Treatment Plant Operation*, a training manual prepared by California Department of Health Sciences and the U.S. Environmental Protection Agency. A similar table is also available for groundwater systems. It is noted that this

table only summarizes applicable treatment processes, and the minimum acceptable treatment must comply with applicable regulations as detailed in further sections.

Table 5.4 Ambient Water Quality Indicators	
Ambient Water Quality Indicator	Treatment Alternatives
Coliforms or microbial contamination	<ul style="list-style-type: none"> Disinfection (chlorination, ozonation, chlorine dioxide, Ultraviolet Irradiation, etc.) Membrane filtration
Turbidity, colour	<ul style="list-style-type: none"> Full conventional treatment (coagulation/flocculation, sedimentation, filtration) Direct filtration (coagulation/flocculation, filtration) if turbidity is consistently less than 5 NTU Membrane filtration
Odours (organic material)	<ul style="list-style-type: none"> Clarification - full conventional treatment Oxidation – Chlorination or Permanganate Special Oxidation – Chlorine Dioxide Adsorption – (Activated Carbon)
Iron and/or Manganese	<ul style="list-style-type: none"> Chemical Sequestration Special Ion Exchange Permanganate and Greensand Oxidation (Aeration, chlorine, permanganate)
Excessive Hardness (Calcium and manganese)	<ul style="list-style-type: none"> Ion Exchange Softening Lime (& Soda) Softening
Dissolved Minerals (High Total Dissolved Solids)	<ul style="list-style-type: none"> Ion Exchange Reverse Osmosis
Corrosivity (low pH)	<ul style="list-style-type: none"> pH adjustment with Chemicals Corrosion Inhibitor Addition

b. Wastewater Collection, Pumping, Treatment & Disposal

Effluent requirements imposed on wastewater treatment facilities are dependent on the characteristics of the receiving body. The minimum acceptable level of water quality is represented by Provincial Water Quality Objectives (PWQO). This defined level of water quality provides a baseline for assessing the quality of the waters of the Province, and acts as a simple measure of ecosystem health. PWQO are typically used as a starting point to develop wastewater effluent discharge requirements for Certificate of Approvals.

Procedure B-1-1: "Water Management – Guidelines and Procedures of the Ministry of Environment and Energy" identifies policies to deal with two situations: 1) where water quality is better than the Provincial Water Quality Objectives; and 2) where water quality presently does not meet the Objectives. Typically, effluent discharge from wastewater treatment plants is allowed in Policy 1 and 2 receivers only, as per the said policy. The two policies are defined as follows:

Policy 1: *"In areas which have water quality better than the Provincial Water Quality Objectives, water quality shall be maintained at or above the Objectives".*

Policy 2: *"Water quality which presently does not meet the Provincial Water Quality Objectives shall not be degraded further and all practical measures shall be taken to upgrade the water quality to the Objectives".*

Site-specific receiving water assessments are conducted to assess the existing conditions, and to determine the assimilative capacity of the receiving body. For river discharges, consideration should be given to background water quality, temperature, flow variations, and downstream water uses. For lake discharges, consideration should be given to outfall depth, currents, thermal stratification, bottom characteristics, and nearby water uses.

Typical wastewater treatment requirements based on the characteristics of the receiving body are described in Table 5.5.

Receiving Body	Typical Treatment Requirements
Great Lakes	<ul style="list-style-type: none"> • Secondary Treatment or Equivalent • Phosphorous Removal • Seasonal Nitrification
Larger River (Not Policy 2)	<ul style="list-style-type: none"> • Tertiary Treatment or Equivalent • Non-toxic Effluent • Phosphorous Removal • Nitrification
Sensitive Receiver	<ul style="list-style-type: none"> • MOE dictates "Best Available Technology Economically Achievable" (BATEA) • Non-toxic Effluent • BOD/ Suspended Solids less than 10 mg/L • Phosphorous less than 1 mg/L • Nitrification

5.3.7 Regulatory Requirements

a. Water Supply, Treatment, Pumping & Distribution

In August 2000, the Ontario government promulgated Ontario Regulation 459/00 Drinking Water Protection Regulation made under the Ontario Water Resources Act. The Regulation provides a legal basis for the Ministry of Environment to regulate the treatment provided by public water systems and includes the Ontario Drinking Water Standards, formerly the Ontario Drinking Water Objectives, which have been updated and strengthened to reflect more current expertise and procedures in drinking water treatment.

Currently in Ontario, standards are set for 99 chemical/physical parameters including turbidity, inorganic constituents, organic constituents, pesticides and PCBs, 4 microorganisms or indicator organisms and 78 radionuclides. Maximum allowable concentrations have been set for 126 contaminants and 22 contaminants have been assigned interim maximum allowable concentrations until sufficient toxicological data has been established. In addition, aesthetic objectives have been set for 27 substances that may impart taste, odour, or colour, and operational guidelines have been set for 5 parameters that require control to ensure efficient and effective treatment.

The new DWPR identifies three different source water types – *surface water*, *groundwater under the direct influence of surface water* and *groundwater* – and mandates the provision of varying levels of treatment for each type.

The Regulation mandates that systems which obtain water from a surface water supply are required to provide a minimum level of treatment consisting of chemically assisted filtration and disinfection

as detailed in Procedure B13-3 Chlorination of Potable Water Supplies in Ontario or other MOE approved treatment, to achieve a minimum 3.0-log (99.9%) reduction of *G. lamblia* cysts and 4.0-log (99.99%) reduction of viruses, based on the characteristics of the source water. This standard is based on the United States Environmental Protection Agency (US EPA) Surface Water Treatment Rule, promulgated in June 1989 but does not take into account the recent revisions to the Rule which now establishes standards for *Cryptosporidium* cysts, through the Enhanced Surface Water Treatment Rule and the Long-Term Surface Water Treatment Rules (LT1SWTR and LT2SWTR).

Because the Regulation considers groundwater supplies under the direct influence of surface water as surface water sources, the level of treatment mandated in Procedure B13-3 for surface water sources also applies to such groundwater supplies, unless it can be shown that the source water quality conditions as well as watershed control strategies are adequate to avoid filtration. In the latter situation, the groundwater supply could avoid the chemically assisted filtration requirement and provide the system treatment requirements of at least 3.0-log reduction of *G. lamblia* cysts and at least 4.0-log reduction of viruses through disinfection only.

The minimum treatment level for a groundwater supply is disinfection designed for a minimum free chlorine residual of 0.2 mg/L after 15 minutes t_{10} contact time at maximum flow and before the first consumer.

As indicated previously, additional treatment processes may be required to ensure that the water supply achieves at least the minimum levels of quality indicated by the Maximum Allowable Concentrations established in the Ontario Drinking Water Standard.

b. Wastewater Collection, Pumping, Treatment & Disposal

Effluent requirements are determined under the provisions of Procedure B-1-1: "Water Management – Guidelines and Procedures of the Ministry of Environment and Energy". Effluent requirements are established on a case-by-case basis and are dependent on the characteristics of the receiving water body, and Federal and Provincial effluent regulations and procedures.

The minimum acceptable level of water quality has been defined by the Provincial Water Quality Objectives (PWQO). This defined level of water quality provides a baseline for assessing the quality of the waters of the Province and acts as a simple measure of ecosystem health. PWQO are typically used as a starting point to develop wastewater effluent discharge requirements for Certificate of Approvals.

Guideline F-5: "Determination of Treatment Requirements for Municipal and Private Sewage Treatment Works Discharging to Surface Waters" dictates the type of sewage treatment process required to meet specific effluent quality criteria. The normal level of treatment is based on secondary treatment, or equivalent. Biological processes, including activated sludge variations or lagoon systems, physical-chemical processes, and a combination are all capable of producing the required level of treatment. If the effluent requirement determined by the water quality assessment is more "stringent" than the normal level of treatment as required in the Provincial guideline, then additional treatment processes will be necessary.

5.4 Costing Part B: Cost Estimate Development

5.4.1 Rules of Thumb

Rule of thumb estimates refer to estimates prepared based on the use of commonly accepted guidelines, conventions or standards and are available for some of the more common capital cost items. These 'rules' are generally applicable for order of magnitude estimates of normal work under

reasonable site conditions and should always be treated with caution and checked using other estimating tools where at all possible.

Examples of these 'rules' would include;

Pumping Stations:	\$ per Horse Power (HP) of total installed pump HP.
Storage Reservoirs:	\$ per unit of reservoir capacity
Water Mains:	\$ per unit of linear diameter per unit of installed length
Six-Tenths Rule:	This is a method of estimating facility costs where the cost and capacity of a similar facility is known. The six-tenths factor is to account for the fixed cost associated with any facility. ⁴⁵

As noted above, these 'rules' should be used with caution and the following issues should be taken into consideration:

- There may be confusion with the units of measurement (US gallons or Imperial gallons; centimeters or millimeters; currency conversion (US \$ to CDN \$)).
- The range of values over which the 'rule' is applicable should be checked (the 'rule' probably will not be applicable to small projects).
- The 'rule' may be old and require updating.
- The 'rules' are generally indicative of average conditions and will not be directly applicable to projects with, for example, significant geological conditions (rock), high water table, deep trenches, congested sites or streets, sites containing hazardous wastes or remote sites.

However, notwithstanding the above caveats, rules of thumb can be useful in the preparation of quick, order of magnitude estimates when applied cautiously by experienced estimators.

5.4.2 Cost Curves

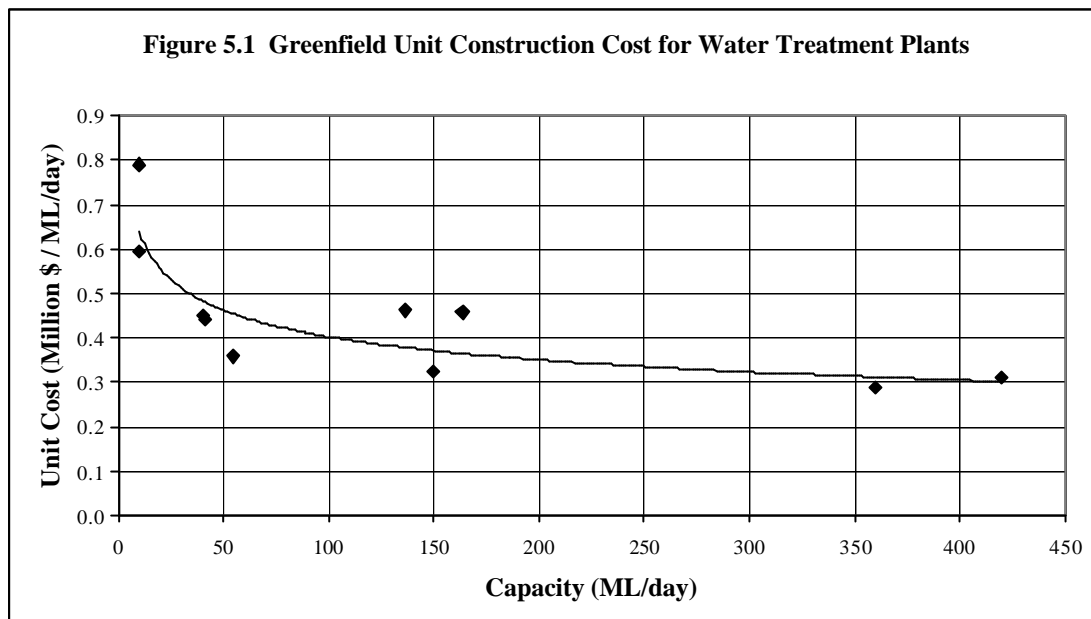
Cost curves have been developed from historical data on capital costs for water and wastewater infrastructure construction. The information used to develop the curves has usually been collected and processed by, or on behalf of, the Federal or Provincial Governments, the US EPA, individual persons or companies through papers published in recognized journals such as the Journal of the American Water Works Association (AWWA), and by Consulting Engineers from their own records and other published information, such as "The Daily Commercial News". When utilizing cost curves it is important that they be keyed to a cost index to allow for adjustment for inflation and geographic location.

Good cost curves are generated from data obtained from large samples where the data has been carefully reviewed to ensure that the costs reflect the total project costs for typical projects in a form suitable for modelling. Information should be available as to what is specifically included in, or excluded from, these costs so that all appropriate factors can be applied to deal with other costs such as inflation, site location, US versus Canadian dollar, contingencies, etc. Clear information allows all costs to be included once, and not repeated.

Curves are available for sanitary sewers, water mains, pumping stations, storage reservoirs, and water and wastewater treatment plants, both for individual process unit costs and for complete systems.

An example of a typical cost curve for water treatment plant construction is shown in Figure 1 below.

⁴⁵ For example, given a conventional water treatment plant with a capacity of a units of capacity and a construction cost of \$ b, the estimated cost of a similar plant of c units of capacity would be $\$b * (c/a)^{.6}$.



5.4.3 Summary of Variables

Regression analysis techniques are used to develop cost curves relating item costs to baseline variables. The baseline variables are dependent on the type of infrastructure to be costed. Common baseline variables include hydraulic capacity, population, unit flow rate and horsepower. Table 5.6 summarizes common cost curve variables for water and wastewater infrastructure construction.

Table 5.6 Cost Curve Variables for Construction		
Description	Baseline	Cost
Water Supply Wells	Well Capacity (ML/d)	\$/ML/d
Water Intakes	Intake Capacity (ML/d)	\$/ML/d
Pumping Stations	Station Capacity (ML/d)	\$/ML/d
	Installed Horsepower (HP)	\$/HP
Water & Wastewater Treatment Plants (Total Cost)	Plant Capacity (ML/d)	\$/ML/d
Water & Wastewater Treatment Plants (Process Unit Cost)	Unit Flow Rate (ML/d)	\$/ML/d (Unit Cost)
Distribution Mains	Diameter (mm) / Length (m)	\$/ linear m (based on diameter)
Sanitary Sewers		
Storm Sewers		
Storage Reservoirs	Reservoir Capacity (ML)	\$/ML
Effluent Outfall	Outfall Capacity (ML/d)	\$/ML/d
Note: ML/d represents megalitres per day (1 megalitre is 1,000 cubic metres)		

5.4.4 Qualifiers

Because cost curves have been derived from past and unspecific information, estimates derived from cost curves, as with all estimating tools, must have qualifications applied to the results to take account of the many variables between the capital works from which the curves were derived and the works to be estimated (Table 5.7).

Table 5.7 Cost Curve Qualifiers		
Description	Qualifications	Notes
Water Supply Wells	<ul style="list-style-type: none"> Unusual geological conditions Excessive well depth Surface protection required Water treatment required 	<ul style="list-style-type: none"> Seek local advice Seek local advice Refer to curves for groundwater treatment Refer to curves for groundwater treatment
Water Intakes (River)	<ul style="list-style-type: none"> Geotechnical constraints (i.e., high rock elevations, erosion, etc.) 	<ul style="list-style-type: none"> Add costs for rock excavation, geotechnical investigations, special foundations, etc.
Pumping Stations	<ul style="list-style-type: none"> Geotechnical constraints (i.e., high rock elevations, poor soil conditions, wetlands, etc.) High water table Additional equipment (i.e., screening, standby power, odour control, instrumentation, VFDs, etc.) 	<ul style="list-style-type: none"> Add costs for rock excavation, geotechnical investigations, special foundations, etc. Add extra for dealing with water Add extra for additional equipment
Water & Wastewater Treatment Plants (Total Cost)	<ul style="list-style-type: none"> Geotechnical constraints (i.e., high rock elevations, poor soil conditions, wetlands, etc.) High water table Presence of hazardous waste Additional or fewer processes 	<ul style="list-style-type: none"> Add costs for rock excavation, geotechnical investigations, special foundations, etc. Add extra for dealing with water Add extra for dealing with waste Refer to curves for unit process costs (add or subtract)
Water & Wastewater Treatment Plants (Process Unit Cost)	<ul style="list-style-type: none"> Additional items above sum of the unit processes 	<ul style="list-style-type: none"> Add percentage for other works (site work, electrical, instrumentation, etc.) based on historic records for percentages of total tender breakdowns. (See Table 4.3 below)
Storage Reservoirs	<ul style="list-style-type: none"> Geotechnical constraints (i.e., high rock elevations, poor soil conditions, wetlands, etc.) High water table Presence of hazardous waste 	<ul style="list-style-type: none"> Add costs for rock excavation, geotechnical investigations, special foundations, etc. Add extra for dealing with water Add extra for dealing with waste
Effluent Outfalls (River)	<ul style="list-style-type: none"> Geotechnical constraints (i.e., high rock elevations, erosion, etc.) Nature of receiving body (i.e., lake, river, stream, ocean) Fresh or salt water Excessive outfall depth or length 	<ul style="list-style-type: none"> Add costs for rock excavation, geotechnical investigations, special foundations, etc. Refer to curves for specific receiving body. Add extra for dealing with salt water Add extra for dealing with depth or length

Cost curves for lake intakes and effluent outfalls are seldom used. With relatively few intakes and outfalls constructed, and each designed specifically for its location, the processing of data to generate curves is very difficult. Factors to be considered in estimating costs for lake intakes and effluent outfalls include length of pipe (which is generally related to water depth) and type of

construction (in-tunnel, in-trench, on lake bed). Estimates for these items are usually made using local information from in-house sources, utility owners, and specialist contractors.

Where estimates are derived from individual Process Unit Costs, other plant costs may be estimated from percentages calculated from tender breakdown information. A typical example is given in Table 5.8 (Source: Peters & Timmerhaus. *Plant Design and Economics for Chemical Engineers*).

Table 5.8 Percentage of Total Facility Cost Based on Historical Data		
Description of Cost	Percentage of Cost	Simple Example (\$)
Equipment Cost	100%	1000
Cost as a Percentage of Equipment Cost:		
Equipment Installation	50%	500
Process Mechanical Piping	65%	650
Instrumentation & Control	20%	200
Electrical	10%	100
Buildings	20%	200
Yard Improvements	10%	100
Service Facilities	70%	700
Engineering and Supervision ¹	35%	350
Project Management and Overhead	40%	400
Total Percentage of Equipment Cost	420%	4200
Subtotal Percentage of Project Cost for Above	80%	
Additional Project Cost Elements		
Misc. & Unidentified Equipment	10%	100
Misc. & Unidentified Process Mechanical	5%	50
Misc. & Unidentified Electrical/ I&C	5%	50
Percent of Total Project Cost	100%	4400

Note: (1) This would be the contractor's cost, not the consulting engineer's design and site review costs. This illustrates how care must be taken in using information from sources where the terminologies may appear ambiguous to the user who may not have access to full details of the sources of the information or how it has been processed.

5.4.5 Other Factors

Cost curves provide a means of generating construction costs for a project. However, to ensure the accuracy of the cost estimate, other factors must also be accounted for (Table 5.9).

Table 5.9: Other Factors Impacting Cost Curves	
Item	Action
Remote Site	Add factor (e.g. 'City Cost Index' as published by R.S. Means Inc.)
Updating estimate from cost curve	Add factor (e.g. Construction Cost Index as published by Engineering News Record (ENR) for US, Southam Construction Cost Index for Canada.
\$(US) to \$(CAN)	For \$(US) cost curves use 'City Cost Index' as published by R.S. Means Inc. This adjusts \$(US) to \$(CAN) for various Canadian Cities. If using this, do not apply 'City Cost Index' again for a remote site as it is already applied here.
Contingencies	Add contingency, as noted elsewhere in this report, appropriate to the level of the cost estimate.
Taxes	Curves based on US projects should be amended to include Provincial Sales Tax on Materials. Federal Goods & Services Tax is generally omitted from estimates.

Table 5.9: Other Factors Impacting Cost Curves

Item	Action
Cost Escalation	Generally escalate final estimate to mid point of the construction period at the, then current, inflation rate.

5.4.6 Sources of Cost Curves

a. In-House

In house cost curves have been, and are continuously being, developed and monitored by Consulting Engineering Companies and their associates from their own records, their Clients' records and from other published information, such as records of successful tenders from "The Daily Commercial News".

b. Literature Reviews

Information is available from cost curves in many of the textbooks in general use by the Engineering Profession in the design of water and wastewater treatment systems, in government reports and in technical papers.

References for commonly used cost curves for water works construction are listed below.

- Abelin, S.M., M.T. Pritchard, and R.L. Sanks, "Chapter 29 – Costs". *Pumping Station Design, Second Edition*. Butterworth Heinemann Publication, 1998.

The data for the cost curves was obtained from a survey of approximately 30 sources in the United States, representing water and wastewater pumping station construction costs from 1966 to 1987. Over 90% of the data is representative of construction costs later than 1974. Data for atypical pumping stations (such as those excessively deep or built to unusual specifications) were either discarded or explained in the text.

- AWWA/ASCE, "Chapter 26 – Construction Costs". *Water Treatment Plant Design, Third Edition*. McGraw-Hill Co. Ltd., 1990.

This publication contains no cost curves but has useful information on special considerations. A table is provided to assist in applying a designated percentage factor to the known cost of the major process equipment for a project, thus allowing an estimate to be made of all other cost factors of the project.

- Cotton, C.A., D.M. Own, G.C. Cline and T.P. Brodeur. "UV Disinfection Costs for Inactivating *Cryptosporidium*". *Journal of the America Works Association (AWWA) (Management Operations)*, June 2001.

This paper provides cost curves developed from estimated costs for a series of conceptual designs for retrofitting UV disinfection into an existing treatment plant. The estimates were made using manufacturers' estimates for equipment with engineering estimates and judgment being used to estimate other associated costs. The reference to this paper is included to indicate how costing information can be obtained from technical papers for newer processes where limited historic data is available and where costs are changing rapidly as these newer technologies are adopted.

- *Drinking Water Infrastructure Needs Survey – Modelling the Cost of Infrastructure: 1999*. U.S. EPA, Washington, D.C., 2001.

The data for the cost curves was obtained from responses to questionnaires sent out to all 1,111 of the largest water system operators, and to a random sample of 2,556 of the 7,759 medium water system operators across the United States. The percent of respondents was 96% and 100% for medium and large system operators, respectively. In addition, random site visits were made to small community systems to gather data. Costs submitted on the questionnaire were subjected to careful review before being included in the data bank to ensure that the costs reflected total project costs for typical projects in a form suitable for modeling. The costs were adjusted to January 1999 dollars. Due to the response to the questionnaires and careful review of the responses, the curves are likely to have been based on reliable, up-to-date data and should provide a useful tool for the preparation of order of magnitude and preliminary estimates. Where no in-house or local cost curves are available, or as a check when using other curves, these cost curves are recommended with the appropriate factors applied for other costs.

- James M. Montgomery, Consulting Engineers, Inc., "Chapter 25 – Construction and Operating Cost Estimating". *Water Treatment – Principles and Design*. John Wiley & Sons, New York, 1985.

Provides some examples of unit process cost curves for water treatment plants developed from historic cost data and detailed engineering estimates obtained from the US EPA study in 1978. Includes excavation and site work for the processes, equipment, concrete and steel, labour, piping, electrical and instrumentation, and housing together with costs of sub-contractor overhead and profit, and 15% allowance for contingencies. The text also provides a curve for engineering costs (ASCE, 1981); curves for interest during construction (U.S. EPA, 1978); curve for legal, fiscal and administrative costs (U.S. EPA, 1978); and a table for general contractors overhead and profit percentage (U.S. EPA, 1978).

- Reid Crowther Ltd. *Canadian Water Treatment Study*. Public Works Government Services of Canada, 2000.

The data used to generate the curves was collected from Canadian sources, including Reid Crowther and MacViro Reports, members of the subcommittee and various government contacts, the Canadian Water Works Association, individual municipal water utilities across Canada, chemical suppliers, transportation companies, the Manitoba Water Services Board and the Saskatchewan Board. The greatest amount of data was obtained from British Columbia, Saskatchewan, Manitoba and Quebec. Limitations in the data were acknowledged in that consistent information was difficult to obtain and it was difficult to determine exactly what was included. The uncertainty range of plus 50% to minus 25% was considered to be appropriate.

- Syed R. Qasim, Siang W. Lim, Edward M. Motley and Kim G. Heung, *Estimating Costs for Treatment Plant Construction*. Journal of the American Water Works Association (AWWA) (Management and Operations), August 1992.

This paper provides curves in the form of equations developed from the 1979 report of the U.S. EPA for 99 unit processes useful in removing contaminants included in the Nation Interim Primary Drinking Water Regulations. The generalized construction costs represented by the equations include excavation and site work for the processes, equipment, concrete and steel, labour, piping, electrical and instrumentation, and housing together with costs of sub-contractor overhead and profit, and 15% allowance for contingencies. The paper also addresses other costs related to the total project cost, and describes indices commonly used to adjust the estimates for site location and inflation purposes.

References for commonly used cost curves for wastewater construction are listed below.

- Abelin, S.M., M.T. Pritchard, and R.L. Sanks, "Chapter 29 – Costs". *Pumping Station Design, Second Edition*. Butterworth Heinemann Publication, 1998.

The data for the cost curves was obtained from a survey of approximately 30 sources in the United States, representing water and wastewater pumping station construction costs from 1966 to 1987. Over 90% of the data is representative of construction costs later than 1974. Data for atypical pumping stations (such as those excessively deep or built to unusual specifications) were either discarded or explained in the text.

- *Analysis of Construction Cost Experience for Wastewater Treatment Plants: Technical Report.* U.S. EPA, Washington, D.C., 1976.
- Bradley, R.M., M.G. Powell, and M.R. Soulsby. "Quantifying Variations in Project-Cost Estimates". *Journal of Management Engineering*, Vol. 6, No. 1, pp. 99-106, January 1990.

The paper describes the method of cost estimating selected for a major wastewater project in Kuala Lumpur, Malaysia. The component cost ranges and probability values were defined by a panel of estimators in order to reduce personal bias. The alternative strategy costs were simulated on a random number basis, the values for each component being selected randomly within the specified ranges.

- Dames & Moore Water Pollution Control Engineering Services. *Construction Costs for Municipal Wastewater Conveyance Systems: 1973-1979.* U.S. EPA, 20460, Washington, D.C., 1981.

The cost curves were developed from a study of construction bid data for wastewater conveyance systems obtained from the ten EPA regional offices across the United States. Approximately 455 construction projects were analyzed, including new and upgraded conveyance facilities.

- Dames & Moore Water Pollution Control Engineering Services. *Construction Costs for Municipal Wastewater Treatment Plants: 1973-1978.* U.S. EPA, 20460, Washington, D.C., 1980.

The cost curves were developed from a study of construction bid data for wastewater treatment plants across the United States. Approximately 536 construction projects were analyzed, including new and upgraded treatment facilities. Ineligible treatment costs and the cost of facilities for collecting and pumping wastewater were not included. Curves are available for three levels of details: the cost of the entire plant, the sum of the costs of the individual unit processes, and the sum of the costs of individual construction components.

- Dayton & Knight Ltd. *Municipal Sewage Discharge Criteria Technical Report No.1.* British Columbia Ministry of Environment, Lands and Parks, 1993.

The cost curves presented were generated from data collected by Dayton and Knight for the Provincial Government in the mid-1990's. Construction costs were obtained for over 30 wastewater treatment plants in British Columbia, Alberta and Washington State.

- Neely, E.S. Jr. and R. Neathammer. "Life-Cycle Maintenance Costs by Facility Use". *Journal of Construction Engineering and Management.* Vol. 117, No. 2, pp. 310-320, June 1991.

The paper details the contents and application of several databases developed by the U.S. Army Construction Engineering Research Laboratory for determining life-cycle costs of building facilities.

- Patterson, W.L., and R.F. Banker. *Estimating Costs and Manpower Requirements for Conventional Wastewater Treatment Facilities.* U.S. EPA, 17080 DAN, Washington, D.C., 1971.

- *Process Plant Construction Estimating Standards*, Vols. I, II, III and IV. Richardson Engineering Services Inc., San Marcos, CA (updated annually).

5.4.7 Software

Computer software, such as CapdetWorks or BACPAC, can also be used to produce budgetary estimates.

CapdetWorks software, available from Hydromantis Inc., is a useful tool in developing cost estimates for wastewater treatment plant construction. The user is required to define the influent and desired effluent quality, and CapdetWorks will calculate capital, operating and maintenance costs for each alternative. Costs such as labour, amortization, land, concrete, pumps, pipes, power, chemicals, and design fees are included. Default costs are included, however all costs can be localized for a specific geographic region, or the user can create their own cost index or apply published industry cost indices.

BACPAC, a proprietary estimating and scheduling system from Brown and Caldwell (1984), utilizes a system cost database, encompassing over 17,000 items for wastewater infrastructure construction to estimate construction costs. Estimates are prepared using labour-hours, local wage rates, and local material costs.

5.4.8 Summary

The accuracy of cost estimates is dependent on the amount of data available to characterize the site and the project requirements. At a master planning level, study or order of magnitude cost estimates are typically used. The anticipated accuracy associated with this type of estimate is plus 50% or minus 30%. Common methods used to generate estimates include standard cost factors and curves, rules of thumb, and computer software. It is important that sufficient information regarding the source of curves and other data be available to so that all appropriate factors can be applied to deal with other costs such as inflation, site location, currency (\$US or \$CDN), contingencies, etc. Clear information allows all costs to be included once, and not repeated.

Additional costs that must be included when preparing a cost estimate include engineering fees, administration and legal costs, land acquisition costs, and provincial and federal taxes. The allowance for these items is dependent on the size and complexity of the project.

Study or order of magnitude estimates are used to compare the economics of various treatment options or the costs of major process components. It is important to realize that this type of estimate does not represent the actual construction and operation and maintenance costs of the project. Actual project costs are site-specific, cannot be generalized, and must be developed based on specific project requirements.

5.5 Econometric Studies of Water Utility Costs

This section has two focal points: first, it describes the major empirical, econometric work carried out on service costs; and second, it provides some non-empirical conjecture about the long run cost function for individual utilities. The empirical analysis focuses on the different scale and scope economies at the plant level, rather than on possible economies of scale involved in multi-plant operation. Multi-plant economies of scale are an important scale consideration in the predominately

government owned utilities operating in large urban centres, or large industries, which provide service from several plants, rather than one single plant. The method of analysis used is normally cross-sectional in nature and generally uses both capital and operating costs to estimate the cost function for the utility. The use of cross-sectional data and capital as a factor input generate a long run cost function for water supply. Earlier studies generally use the Cobb-Douglas functional form while later studies use the trans-log form.

Some of the earlier studies focus on different components of the water supply system, with other studies examining the costs of the entire system. The following subsections will follow rough chronological order in describing the different studies by system component, with the first section describing studies of separate system components (e.g., treatment plants), the second total water supply system costs, and the third wastewater treatment system costs. The section will make some conjecture on the shape of the cost function for individual utilities with some variables not considered in the empirical work (e.g. source of supply, multi-plant economies).

5.5.1 Empirical Studies of Water Supply Component Costs

Orlob and Lindorf (1958) carried out an early empirical study of water treatment costs, using data from 32 California treatment plants. The technology then, as now, was flocculation, sedimentation, filtration, and chlorination.

Though no statistical testing was done on the data, graphical analysis of engineering data enabled Orlob and Lindorf to estimate capital and operating cost functions. The capital cost function estimated was:

$$(5.1) \quad C_c = 257Q^{0.67}$$

where: C_c = capital cost; and
 Q = design capacity.

The operating cost function was:

$$(5.2) \quad C_o = 68.4Q^{0.59}$$

where: C_o = operating cost.

The results for the scale parameter in the capital cost function were confirmed by Koenig (1965) in a study using OLS and a logarithmic transformation of the independent and dependent variables, on a sample of 30 treatment plants, again using engineering data, where the capital cost function derived is;

$$(5.3) \quad C_c = 30.7Q^{0.677}$$

with a R^2 of 0.59 and a significance level of 0.001⁴⁶.

These results reflect the widespread use of 2/3 rule in engineering, where for some production processes output is in rough proportion to the volume, and cost in rough proportion to the surface area. Because the area varies in 2/3 proportion to the volume (for a sphere or cylinder), costs will rise in a 2/3 proportion to design capacity. According to Moore (1959) these results apply specifically to industries with the following characteristics:

⁴⁶ Unless stated otherwise all significance results refer to a regression coefficient >0.

- The production process is continuous and not batch oriented,
- The industry is capital intensive, and
- The industry produces a homogeneous product,

Though Koenig did not directly derive an operating cost function, Hiromoto (1971), using Koenig's data, used a logarithmic transformation and OLS to derive separate operating cost functions for chemicals, pumping energy, heating energy, manpower, maintenance and repair as well as capital investment, given an arbitrary use rate of 50% of design capacity. All of the operating cost functions developed indicate economies related to operating costs and capacity, with regression coefficients of 0.77, 0.72, 0.47, 0.69, and 0.58 respectively for chemicals, pumping energy, heating energy, manpower, and maintenance and repair.

The study by Hiromoto was the latest to analyze water treatment cost alone. Subsequent studies that examined the combined affects of both treatment plant costs and distribution costs will be discussed in the sub-section on total costs.

There is only one empirical study that attempts to do full justice to the problems involved in measuring the economies of scale involved solely in the transmission and distribution system (Linaweaver and Clark, 1964), a study incorporating both capital costs and the operating costs (primarily energy) involved in delivering water to each connection.

The capital costs were estimated using the pipe diameter as a measurement of pipe size such that the relationship between capital costs and pipe diameter (inches) was:

$$(5.4) \quad C_c = 2.16D^{1.2}$$

where: D = pipe diameter (in inches).

Though this equation seems to indicate diseconomies in scale related to pipe diameter, it is questionable whether the dependent variable is chosen correctly. The appropriate variable, it is suggested, might more appropriately be pipe capacity rather than diameter, and because pipe is cylindrical, the 2/3 engineering rule would appear to apply.

Scarato (1969) used the 2/3 rule to estimate capital costs for pipe in the distribution network where the capital cost function is hypothesized to be;

$$(5.5) \quad C_c = AQ^{0.67}$$

where: Q = demand capacity; and
A = a constant.

Two other problems associated with the Linaweaver and Clark (1964) study were the exclusion of capital costs related to the pumping stations and storage facilities located in the distribution system, and the exclusion of a length variable indicating the total length of pipe in the distribution system. Following Goddard, Stevie and Trygg (1978) used a hypothetical capital cost function to estimate capital cost; the Cobb-Douglas form of this function was:

$$(5.6) \quad C_c = AQ^a L^b$$

where: Q = demand capacity (maximum daily or hourly flow, depending on storage);
A = a constant;
L = the length of pipeline; and
a and b are scale parameters.

The operating cost formula for distribution is difficult to calculate because the energy required to pump the water through the network is dependent on the total dynamic head in the system. Total dynamic head consists of three components: static head, velocity head, and friction head, where head can be defined as pressure (measured in pounds per square inch, or kilos per sq. meter) Static head refers to the initial pressure produced by elevation, velocity head refers to the pressure needed to force water through a pipe at a given velocity, and friction head refers valves and fittings of the system.

The methodology for determining the total head needed to transport water is an engineering subject that has generated a considerable engineering literature, the survey of which is beyond the scope of this paper. A discussion of some of the complications can be found in Goddard, Stevie and Trygg (1978).

As an illustration of the complexity of water system cost functions, Linaweaver and Clark defined operating costs to be energy costs plus 8% (to account for operating and maintenance costs) where average energy costs are:

$$(5.7) \quad AC_e = P(1.66 \times 10^{-2}) [.75L + .667(10^3 Q^{1.85}) / (405 \times 10^6 CD^{2.63})^{1.85}] / E$$

where: P = pressure (feet per square inch),
 Q = flow capacity (millions of gallon per day),
 L = the slope of the system (feet),
 C = a roughness coefficient,
 D = distance (miles), and
 E = energy, based on a capacity utilization rate of .667.

The results indicated decreasing returns for operating costs in the distribution system as both capacity and pipe length increase.

Based on the Clark and Linaweaver study, as well as the theoretical work of Coase (1947), it is probable that the returns of operating costs in the transmission and distribution network are decreasing as distance or area served by the system increases. As a result different marginal costs are generated throughout the distribution network for each connection served by the utility.

These results also give rise to the observation that economies in production can be offset by diseconomies in distribution, a major theme of the empirical analysis on total system costs using actual utility data, the subject of the following section.

5.5.2 Empirical Studies of Total Supply System Cost

A number of analyses have examined total system costs; these studies generally use more advance theoretical work, and are based on actual utility data. Much of this work is aimed at demonstrating that economies of scale for water utilities were not unlimited.

Ford and Warford (1969) initiated this line of research in a study of the effect of a spatial variable on total system costs in the U.K. This study uses cross-sectional utility data on 331 utilities. Ford and Warford estimated a total cost function using several different forms, with the best fit being obtained with OLS and a logarithmic transformation of the dependent and independent variables. The dependent variable was total system average cost as measured by cash based accounting, with the independent variables being quantity supplied (daily, thousand gallons) and area served by the water utility (square miles). The cost function derived from the modeling is represented by;

$$(5.8) \quad \log AC = 3.78 + .133 \log A - .124 \log Q$$

where: AC = average cost;
A = area; and
Q = quantity.

The resulting function (equation 5.8) had a relatively low R^2 value of 0.22 with each independent variable significant at the 5% level. It is possible that the low R^2 value reflects the use of cash based accounting, whereby total cost may vary depending on the means of financing used. The result indicates increasing return to scale in terms of water supplied, and decreasing returns in terms of area served, implying an optimal size for water utilities.

Andrews (1971) conducted a similar study using the same functional form but somewhat different variables for New England water utilities, where the dependent variable chosen was average cost based on total revenues, under the assumption that total revenues equal total cost. The independent variables used were the number of connections, and quantity supplied (million gallons per day).

The estimated function for average cost was;

$$(5.9) \quad AC = .05Q^{-.7}CO^{.75}$$

where: AC = average cost;
CO = the number of connections; and
Q = quantity.

The modeled function had an R^2 of 0.43, with both variables significant at the 5 % level. This indicates the same results as the Ford and Warford study, in which average costs decreased in terms of quantity served, and increasing in terms of the number of connections, again indicating an optimal size for water utilities.

Comparing the two studies raises the question of which is the correct spatial variable (i.e. area, connections, or density) to use in estimating optimal system size. Clark and Stevie (1981) suggested a possible response to this problem - an engineering study using three spatial variables to estimate average costs under the assumptions of a uniform circular service area, and a constantly diminishing population density moving from the centre to the perimeter. The three spatial variables used were population density at the centre of the urban core, the rate at which density diminishes approaching the perimeter (i.e., density decline per mile), and the distance from the centre to the perimeter.

Clark and Stevie maintained that the most important variable in determining optimal system size is the rate at which density declines up to the perimeter, rather than distance (or area), or density (or number of connections). This would seem to make sense in terms of transmission and distribution cost, particularly capital cost, for as the density declines in moving to the perimeter, the pipe length needed to supply each connection also increases, generating decreasing returns to scale for capital investment in terms of the number of connections, as well as operating costs in the transmission and distribution system.

Hines (1969) was the first to include capacity utilization⁴⁷ as a major determinant of average system costs, using data from 11 Wisconsin utilities, with the independent variables being original indexed plant cost (or historical plant value), and the rate of capacity utilization. The average cost function estimated was:

$$(5.10) \quad AC = 328 - 1.8CU - .000009HPC$$

⁴⁷ Capacity utilization typically refers to the peaking characteristic used to size plant capacity, maximum daily demand.

where: AC = average cost,
CU = the rate of capacity utilization, and
HPC = historical plant cost.

This function indicates increasing returns relative to both the rate of capacity utilization, and the historical plant value (presumably due to its correlation to plant capacity). The methodology used was a simple linear regression without the common logarithmic transform. Unfortunately no R^2 or significance results were reported.

A relatively modern study that uses both the rate of capacity utilization, and the area served by water utility in order to determine a cost function is a study by Kim and Clark (1988). This study uses the translog production function, and the "iterative Zellner estimation method in order to generate maximum likelihood estimates" of the parameters for the translog function. The study also adds the variable of product mix to explain both economies of scale and scope in water production, by defining the utility as a multi-product firm providing a water supply service to both residential and industrial users. The multi-product nature of the service arises from the differing distribution systems needed to provide potable water to both industrial and residential connections.

The dependent variable used was total cost, while the independent input variables used in addition to capacity utilization, distance from plant to perimeter, residential and industrial production are the costs of labour, capital and energy. The database is 63 perpetual public utilities in the United States.

The study found that economies of scope exist in providing both residential and industrial services, but that the product specific economies of scale for residential service are increasing, while those for industry are declining, in each case holding the production of the other service constant. The overall scale economies (i.e. both services together) suggest the relationship found in the other studies where the overall cost elasticity for the multi-product plant is represented by;

$$(5.11) \quad E=0.79 - 0.32\ln YR + 0.25\ln YN + 0.07\ln WL - 0.09\ln WK + 0.03\ln WE - 0.461\ln ZU + 0.11\ln ZD$$

where: E = cost elasticity of output;
YR = water production for residential use;
YN = water production for industrial use;
WL = labour costs;
WK = capital cost;
WE = energy cost;
ZU = rate of capacity utilization; and
ZD = distance from the plant.

Because the overall scale elasticity is given by:

$$(5.12) \quad SL=1/E$$

where: SL = overall scale elasticity; $SL > 1$ indicates economies of scale, $SL = 1$ constant returns to scale and $SL < 1$ diseconomies of scale.

The study found that increases in the rate of capacity utilization and residential production generate increases in scale economies, while increases in industrial production and distance reduce scale economies. The study also examined the optimal plant size given the differing effects of production and distance served. Given the data, the optimal plant size for this study was found to be 22 million gallons per day, serving a distance of 448 miles.

The study concluded that, because the water industry is not a natural monopoly (according to the definition of sub-additivity), the only reason that the water industry "behaves as a natural monopoly" is the influence of regulation, implying that "the market can accommodate more than one firm in the water supply industry"⁴⁸. This conclusion may be seen as referring specifically to separate systems for industrial supply, where specific industries or industrial parks may find it more cost-effective to self-supply rather than hook up to the local municipal or regional system. This is a relatively common occurrence, especially in smaller less urbanized regions.

A second interesting study by Kim (1995) examined the prices of water utilities relative to marginal cost and second best pricing (e.g. Ramsey pricing). The water utility sample and econometric methodology follows the paper listed above. Marginal costs are estimated and compared with hypothetical demand elasticities. Though the price structure deviates from marginal cost pricing, it is similar to second best price discriminations among user classes, in this case residential and commercial classes. This study indicated that the engineering rate setting methods as recommended by the AWWA comprised a sub-set of several second, or first best efficient, solutions possible to pricing water utilities. The principal theories and practices of water utility pricing can be found in the next chapter.

5.5.3 Economies of Scale and Scope in Waste Water Treatment⁴⁹

A smaller number of studies have addressed the issue of sewage treatment costs. For example, Hanke and Wentworth (1981) employed engineering data on construction and operating costs of representative sewage treatment facilities. The authors regressed total cost on output and output squared and find support for the presence of scale economies, similarly to water supply. Fraas and Munley (1984) estimated separate Cobb-Douglas regressions for capital and operating costs using data from a sample of American sewage treatment facilities. Fraas and Munley found that increases in both flow rates and the concentration of sewage raise costs and that the marginal cost of pollution removal increased at an increasing rate with the percentage of pollutant removed. Neither of these studies includes input prices as explanatory variables in their cost functions.

5.5.4 Economies of Scale and Scope in Utilities Comprising Both Water Supply and Sewage Services

A recent study of total utility costs, including water and sewer services, in the province of Ontario was carried out by (Renzetti, 1999). Estimation of the costs of water supply and sewage treatment was carried out separately, under the assumption that municipal utilities seek to minimize the costs of supplying exogenously determined quantities of output. In their choices regarding input use, water supply and sewage treatment utilities are constrained by exogenously determined market prices, their production technologies and the characteristics of their operating environments. These assumptions imply that the two technologies may be represented by their respective cost functions:

(5.13)

⁴⁸ Teeples and Glyer empirically test several different forms of cost functions to determine the effect of ownership on cost efficiency. Their results indicate that the most completely specified cost function generates no efficiency effects from ownership, with a relationship between the variables (ex. Capacity, population density) similar to the studies discussed above (Teeples and Glyer, 1987).

⁴⁹ This section is based on (Renzetti, 1999).

$$C_w = C_w(p_1, \dots, p_N, Q_R, Q_{NR}, D); \quad C_s = C_s(p_1, \dots, p_N, Q_s, D, Z_1, \dots, Z_F)$$

where the subscript w = water supply and s = sewage treatment.

For both water supply and sewage treatments, cost was measured as the sum of annual expenditures on labour, energy and capital. The data used were cross sectional observations of 77 Ontario municipal utilities operating in 1991. The observations were compiled from several sources including national surveys of water utility operations (Environment Canada Municipal Pricing Survey) and municipal financial records (Financial Information Returns) reported to the Ontario government. The sample was representative of the province in terms of the distribution of utility size although the sample was skewed slightly towards larger utilities; the average annual output for utilities in the sample is 8.1 million cubic metres while the provincial average is 6.9 million cubic metres.

All of the utilities were self-supplied and did not purchase their water from a wholesaler. It was assumed that the price vector in both cost functions contained the same three elements: labour (P_L), energy (P_E) and capital (P_K). In the case of water supply, output is a vector of residential output (Q_R) and non-residential output (Q_{NR}). The non-residential customer class was composed of commercial, industrial and institutional customers. Sewage treatment output (Q_s) was taken as a scalar measure of total recorded annual flow-through. In the case of sewage treatment, Z is a vector composed of a set of dummy variables indicating the type of treatment process employed. Finally, the variable D measures the population density of the municipality.

A translog functional form, with the water supply and sewage treatment cost functions estimated separately, approximated the structure of each cost function. Each cost function and $N-1$ of its share equations were estimated using an iterative, SUR procedure with linear homogeneity in prices and symmetry imposed. Table 5.10 shows the average estimated marginal costs for water supply to residential and non-residential customers and for sewage treatment. It also lists the average values for scale elasticities calculated at the mean of the respective data sets. In the case of water supply, the product-specific scale economy measures were calculated following Kim (1987) in a study mentioned earlier.

The estimated marginal cost reported in Table 5.10 is higher than the estimates reported by other researchers. For example Renzetti (1992b), in a case study of the Vancouver waterworks, reported marginal cost values that range from \$0.53/m³ to \$0.85/m³. In addition, the estimates presented here are greater than those reported by Kim (1987) and Russell and Shin (1996) where LRMC estimates based on U.S. data range from \$0.30/m³ to \$0.56/m³ (when converted to 1991 Canadian dollars). The reasons for this divergence may be the effects on construction and operating costs of the lower average temperatures and greater temperature variability, higher labour and interest costs and the lower population densities of Ontario municipalities compared to U.S. cities.

Table 5.10 also indicates that there were scale economies in the technology of water supply and sewage treatment. This finding corresponds to similar results by other researchers. As indicated in the previous section, for example, Kim (1987) found that scale economies are prevalent but that they decline with the size of the utility. While not shown in Table 5.10, cost function coefficients also provide estimates of the elasticity of cost with respect to density. At the mean of the data set, this parameter has an estimated value for water supply and sewage treatment, respectively, of -0.061 and 0.056, though both results are insignificant at the 95 % level.

Table 5.10 Marginal Cost and Scale Economies for Water Supply and Sewage Treatment	

	Marginal Cost	Scale Economies
Residential Supply	0.873* (0.153)	1.249* (0.149)
Non Residential Supply	1.492* (0.398)	1.465* (0.074)
Sewage Treatment	0.521* (0.148)	1.364* (0.755)
Notes		
<p>1. Marginal cost and scale economies are calculated at the mean values of the data and is measured in \$1991/m³</p> <p>2. A scale measure greater than one indicates increasing returns to that output.</p> <p>3. Figures in parentheses are estimated standard errors. A single asterisk indicates significance at the 0.05 level while a double asterisk indicates significance at the 0.10 level</p>		

The Renzetti study also estimated aggregate residential and non-residential water demand equations. Using these estimated relationships, the study then compared the estimated marginal costs to water supply and sewage treatment prices and calculates the gap between predicted consumption levels and those that would be predicted if utilities followed marginal cost pricing. Finally, the paper estimated the welfare costs of these observed divergences and considers the reasons for these findings. The author stated:

“...the most important finding is that prices understate the marginal costs of providing these services by a wide margin. This situation encourages excessive consumption on the part of households and businesses and over-expansion of water supply and sewage treatment facilities. It also discourages technological innovation in water conservation and alternative sewage treatment technologies.”

Two recent studies focused on economies of scope between sewage and water supply stem from the British privatization of water utilities, where prior to privatization the ten integrated river basin management authorities (National Rivers Authorities, NRA) controlled both regulatory and operating functions for water utilities (including supply and sewage)⁵⁰. At privatization the operating and regulatory functions of the NRA were separated, with operating functions transferred to privatized regional water and sewer companies (WASCs), and regulatory responsibilities to separate offices for drinking water, sewage and economic regulation (e.g., OFWAT). The initial study (Hunt and Lynk, 1995), used a parametric, dynamic, multi-product cost function regressing cash based accounting operating costs on output indicators (e.g. quantity of water supplied, treated, regulatory revenues) of the three functions of the previous NRA (supply, sewerage, regulation), as well as labour costs, and a geographic dummy variable. The data set include 90 observations for the 10 RWAs from 1980 to 1988. The regression results indicated that positive production complementarities existed between water supply and sewage, and, particularly, between regulation and environmental services, while negative production complementarities exist between sewage and environmental regulation.

Saal and Parker (2000) criticized the methodology of the Hunt and Lynk study for its use of accounting data, and not including capital costs. The Saal and Parker study assessed the performance of water utilities, both pre- and post-privatization using utility data obtained from the utilities sale prospectus' as well as standard utility accounting in reporting to OFWAT.⁵¹ The study used a trans-log cost function comprising water and sewer outputs, and input prices for capital, labour, and other operating expenses using the Zellner method of OLS, as in Kim (1987). Both of

⁵⁰ Integrated and regulated within the RWA were 29 private water supply utilities, supplying approximately 23 % of water supply. These utilities have remained private.

⁵¹ Conclusions of the study on the impacts of privatization can be found in Chapter Seven.

the studies adjusted basic output to reflect quality considerations, for example compliance with drinking water quality directives, or surface water quality changes, with the Saal and Parker study running regressions on both quality adjusted and unadjusted models. The authors concluded that the separability of inputs and outputs can be rejected, indicating it is inappropriate to model water utilities without using a multi-output model. However, the hypothesis of non-jointness (no complementarities) could not be rejected, with both coefficients (e.g. quality-adjusted, unadjusted) being insignificant at the 95 % level. However, the sign of the jointness parameter changes from positive to negative in the quality-adjusted case, with a much lower level of insignificance. A second interesting finding was that the cross-output interaction terms for water and sewer outputs become negative, and statistically significant, in the quality-adjusted case. The authors interpreted their empirical results as “ suggesting the possible existence of quality driven scope economies in which the improvement in the quality of one output (e.g. sewage) may reduce the cost of producing another (e.g. water supply)”.

5.5.4 Conjecture on Some Variables not Considered in the Empirical Work

The empirical analyses of water utility costs cited in the last section focus on economies of scale and scope at the plant level. Two additional considerations that may be important in determining the cost characteristics of an individual utility are the effect of technological change, and possible economies of multi-plant operation.

The two major types of technological change that affect system costs are changes in the source of supply, and changes in the treatment method. Changes in the source of supply for a water utility affect the costs of the intake system that supplies raw water to the treatment plant. These changes take place primarily as growing demand exhausts previous supplies, but may also occur due to pollution of existing sources. Some of the most common examples are changes from underground aquifers to surface water sources, and changes from small, relatively unpolluted surface sources (e.g. glacial lakes) to major river systems. Changes in treatment technology (e.g. none, chlorination, ozonation) affect the costs of the treatment plant, and generally occur in order to meet sanitary or public health standards. These changes in technology are often correlated, as the change from a smaller unpolluted source to a larger source will also generally require substantial changes in the treatment technology.

These discrete changes in the cost function due to changing technology can generally be expected to generate decreasing returns, particularly in terms of the initial capital investment. This occurs because utilities can be expected to initially choose the least expensive source, or treatment technology, to meet their needs, and then move to the next least expensive as circumstance change (e.g. growth in demand, changes in health standards).

A final factor that influences utility costs is the multi-plant economies of scale derived from shared administration, billing, planning monitoring and finance, a key consideration for large urban communities, or large industrial firms operating many plants. The economies relating to these factors can generally be expected to be increasing as utilities grow from a single plant operation to a multi-plant operation, due to either growth in demand or the merger of smaller municipal utilities into utilities controlled by regional governments. In terms of water utility companies scale economies may be realized as firms acquire more lease or concession contracts. These economies can be realized due to the sharing of planning, administration, monitoring and billing functions over the different plant service populations, as well as the lower financing rates available to larger, and therefore less risky, levels of government or corporations. At some point of multi-plant growth these economies can also be expected to tail off as the administrative hierarchy becomes too unwieldy. Multi-plant scale economies may be large, judging by the generally highly concentrated markets for lease or concession operating private water companies (e.g. Suez).

5.6 Summary

The results of the empirical studies cited above demonstrate the effects on water supply and sewage costs of four principal variables or types of variables, plant capacity, the rate of capacity utilization, spatial variables (e.g. area, density), and the service mix (e.g. proportion of industrial, residential, and commercial connections).

The results indicate long run increasing returns to scale, long run economies in both capital and operating costs, as well as short run economies in operating costs as capacity utilization increases. The result for capacity utilization reflects the discrete nature of capital investments in plant capacity, where plants are normally built to service demand in any given region for a considerable period of time, possibly 20 to 25 years into the future. Thus at any given time the majority of plants in a cross-sectional analysis will be operating at considerably below capacity, where as demand increases and more capacity is used, increasing returns result. In an individual utility increasing returns can be expected to convert to decreasing returns at or near full capacity utilization (as in standard micro-theory), at which point a new plant would be built for the next 20 to 25 years.

The spatial and service mix variables determine the returns to scale within the distribution system, where population density and area served by the water utility act to determine the length of pipe and energy costs needed to distribute water, while the service type influences pipe capacity per connection. These distribution costs demonstrate decreasing returns as the area supplied by a given plant increases, and as density declines towards the perimeter of the service area. As the service type determines pipe capacity an increasing proportion of residential connections can also be expected to produce decreasing returns in the distribution network.

This trade-off between increasing returns relative to production, and decreasing returns as the distribution network expands determine an optimal system size (or minimum efficient scale). In any given community the optimal size will be determined by the interrelationship between the variables outlined above. Economies of scope in supply appear to exist between services to distinct user classes (residential, industrial) while economies of scope between sewage and water supply are not found in the literature, possibly indicating some scope for disaggregation of typical water utilities. An Ontario study estimating water supply and sewer marginal costs finds that marginal costs in Ontario far exceed prices charged.

Additional considerations in determining the cost function for any given utility are changes in technology (either treatment or source of supply). These changes will serve to generate increasing returns to scale, as lower cost sources would have initially been used. There are also possible multi-plant economies of scale through shared administration and expertise in either large urban areas, or large specialized corporations.

CHAPTER 6: WATER UTILITY PRICING THEORY AND PRACTICES

6.1 Introduction

The preceding chapters have described characteristics of municipal water utilities that are important for economic and financial decision-making. In broad terms, these characteristics have focused on viewing these utilities in a natural resource context, and on the detailed analysis of their demand and supply characteristics. These discussions form the background for this chapter, which deals with the economic theory and practice of rate setting, in both theoretical and practical terms. The latter is important because it provides the basis for setting volumetric water rates in many Ontario water utilities. It is important to note that the chapter does not comprise a rate-making manual per se, for this is beyond the scope of the report. Instead the focus is on the means by which the background of economic principles and research findings can be marshalled in a general sense to design effective water rates.

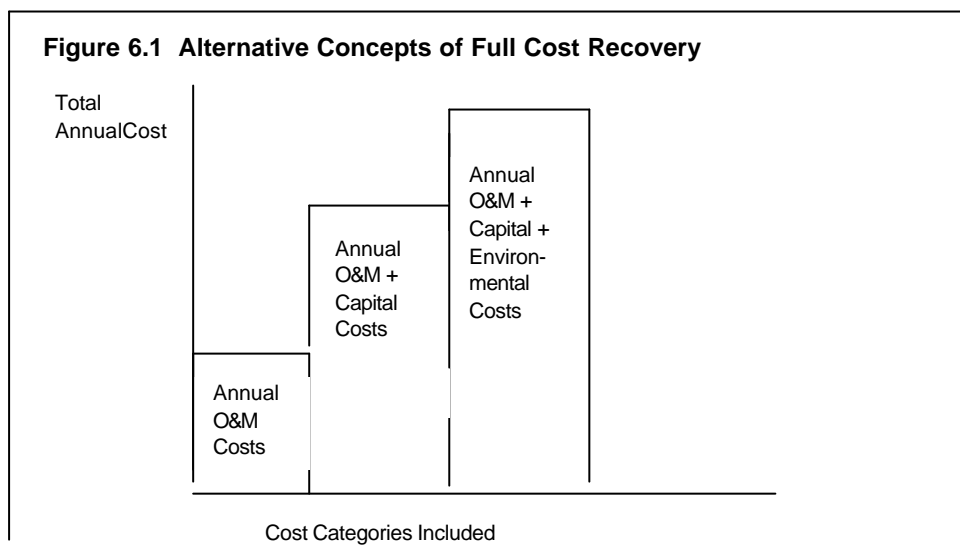
The chapter begins with a discussion of some basic definitional issues associated with an economic interpretation of full cost recovery, which, ideally, water prices in individual municipalities should ensure. This assessment is conducted from an economics perspective, because economic efficiency is viewed as an equally important aim. These two sections lay the groundwork for outlining the economic theory associated with full cost recovery in capital intensive utilities characterized by sub-additivity in the cost function. Again, it should be understood that this discussion is theoretical in nature, focusing (a) on the static and dynamic economic theory of efficient pricing to achieve full cost recovery and (b) on optimal capacity given natural monopoly. This technical economics section (6.3) may be skipped by non-specialist readers. Based on this material, the next section outlines the principal practical methods designed for setting volumetric water rates on a full cost recovery basis in Ontario. These could be used at a later stage to design effective rate structures. This outline is followed by a brief examination of actual pricing practices and levels in Ontario, in order to illustrate the basic underlying problems, such as low prices, poor incentives for effective water use practices, etc. The final section looks at some empirical estimates of the monetary requirements required to achieve full cost recovery in Ontario municipal water servicing. This section has been included to convey a general idea of the “order of magnitude” of the resources required to assure adequate water servicing within the province.

6.2 What Is Full Cost Recovery?

The term “full cost recovery” may mean different things to different people. Broadly speaking, it encompasses at least two broad types of costs – operating and maintenance costs and capital costs⁵². As their names suggest, the former basically involve the annual costs of operating and (routinely) maintaining the utilities. These are quite straightforward to calculate. The latter constitute payments for the “hard assets” of utilities – pipes, pumps, treatment plants, etc. Accounting for these costs is somewhat more difficult, because of variations in accounting and economic practices that can be used. Quite recently, arguments have been put forth to include a third element – the environmental costs, or externalities, associated with water withdrawal and discharge by municipalities. Valuation of these cost is much more difficult and experimental currently, and will not be addressed in this report.

⁵² Raftelis (1989) discussed each of these types of cost in detail.

Conceptually, these three elements can be shown in a simple diagram (Figure 6.1). The levels of effort required to determine these various costs increases from left to right. It is relatively simple to determine annual O&M costs. The issue of capital costs is significantly more difficult, and the bulk of this section is devoted to this issue, though from an economic perspective. For example, one major accounting issue relates to the use of cash based or cost based accounting in taking capital usage into account (a subject addressed in Appendix 6.A). The issues of environmental cost are more difficult still to address, because the theory of environmental damage evaluation is more experimental, and the property rights allocation related to those externalities are primarily the responsibility of higher levels of government.



6.2.1 Economic Prerequisites for Full-Cost Recovery

In economics, the concept of marginal cost pricing is of central importance, due to the efficiency and information provision characteristics of that pricing method, as discussed in chapter two. As part of an economic report, this section assumes that any definition of full cost recovery should incorporate these basic efficiency considerations, through setting utility rates at levels approximating long run marginal costs. This implies incorporating both operating and maintenance costs and capital costs in utility rates. The use of long run marginal cost means that all utility consumers will pay a price that approximates the actual long run costs they impose on the utility for providing that service. This, in turn, will provide accurate information to utility consumers, and provide a basis for making informed decisions regarding water use. While short run marginal cost pricing is in theory, possibly efficient, current preferences for rate and revenue stability in the face of the substantial capital indivisibilities prevalent in water utilities mitigate against it being an acceptable method of full cost pricing, as will be discussed in more detail below in section 6.3.

Accordingly, full cost recovery water rates should incorporate at least three basic efficiency characteristics:

- peak load pricing;
- recovery of marginal distribution network costs through connection charges; and;
- forward looking volumetric rates, incorporating long run marginal cost pricing; and estimates of future, rather than sunk, capital costs.

Possibly a fourth element should also be included, namely extra strength sewer surcharges, which, as their name implies, involves charges on effluent volumes or strengths over and above those that municipal plants are designed to treat. This is primarily an option for municipalities serving a large number of industrial establishments.

a. Peak Load Pricing

Peak load pricing refers to the practice of charging higher prices at times of peak demands. The rationale is that much of the capital expenditure on water utilities stems from meeting peak demands, particularly summer water demands. Peak load pricing is a practical and efficient pricing method, which could be instituted using exactly the same cost allocation methods as the base extra capacity (BEC) method, a common engineering rate setting method recommended in the American Water Works Association (AWWA) rate setting manual. They can also have a significant effect in lowering demands in peak periods, and therefore, in the long run, lowering capital costs.

The BEC method calculates the extra-capacity, or peak, cost and then allocates these costs to different user classes on the equity grounds that low volume users (e.g. residential) have higher peaking demands, and thus should pay a higher share of the capital costs. This results in residential consumers paying higher prices, and industrial consumers lower prices throughout the year. Peak water uses are most likely to occur during summer months, regardless of daily peaking patterns. Therefore, a simpler, and more correct, efficiency-based allocation of peak costs is that all summer water users contribute to the summer peak, and thus all summer water users should be charged for it. The result of this would be lower capacity costs over time, because demand elasticities are higher in the summer, and capacity costs are determined primarily by summer peaking. Peak load pricing is also a better means of water conservation, since it provided incentives for lowering water demands during the period in which water supplies are most likely to be constrained, at least in Ontario (and much of the rest of Canada.)⁵³

b. Connection Charges

Connection charges refer to the costs of installing local distribution systems in new housing or industrial development. They commonly take the form of lot levies, frontage fees, or development charges. Full recovery of these costs from developers and new establishment owners provides correct information on the costs that new development imposes on water distribution systems. Full cost recovery of these charges results in efficient growth in system size, both in terms of limiting urban sprawl, and limiting peripheral connections that are currently self-supplied, such as industry and population located in rural areas. It be recalled from an earlier discussion that decreasing returns to scale occur in water utilities as population density decreases towards the perimeter of the service area.

c. A Long-Run Marginal Cost Perspective

Long run marginal cost pricing in the volumetric rates depends on using estimates of future capital cost to calculate water rates, rather than historical costs, as in the engineering methods described in the AWWA manual. The simple rationale is that historical costs are “sunk costs; or cost which cannot be altered by changing current behaviour. In contrast future capital costs related to system expansion are costs that can be altered by increasing or decreasing water demands, notably by bringing forward or delaying capacity expansion (for a further discussion, see McNeill and Tate, 1990).

d. Extra- Strength Sewer Surcharges (ESSS)

⁵³ It is noteworthy this is not the case on the west coast of Canada, where a “Mediterranean” type of climate tends to concentrate precipitation during the winter months.

Municipal water utilities are designed to meet the needs of residents, plus those of supporting economic activities – commercial, smaller industries, institutions, and the like. Waste treatment systems, accordingly, are sized and designed essentially to treat organic wastes of strengths and volumes generated by these users. A frequent occurrence is that some individual establishments, usually industries⁵⁴, discharge wastes that are excessive in terms of volumes and/or strengths. These impose extra costs on the utility. Theoretically, there is a close relationship between ESSS and the type of economic instrument proposed by Baumol and Oates (1988), as discussed in chapter three.

In many countries, including Canada, ESSS systems are used to levy appropriate charges on excessive industrial wastes (OECD, 1999). Generally, in Canada, this instrument is used by larger municipalities, which treat large volumes of industrial wastewater, although recent data are unavailable as to how many municipalities are involved. Generally the surcharges are based on formulas, which combine two or more pollutants in calculating the number of “units” of polluting material that will be used to calculate the total charge to individual plants where the total charge is equal to the number of units of pollutant(s) times the unit charge.

e. Implications for Full Cost Recovery

Each of these criteria, incorporated in the water rates, will result in utility consumers paying the marginal costs they impose on the system. Customers have a measure of control over their payment of these costs through changing their behaviour, for example in decisions made on reducing or increasing water use, effluent or by connecting or disconnecting from the system. In each instance where utility consumers are charged prices below marginal costs, some degree of cross-subsidization must exist in the pricing system, because the financial costs of water utilities are always eventually recovered from some part of society, be they municipal rate payers, or provincial or federal taxpayers.

However, where prices are higher than marginal costs, it is not certain that cross-subsidization exists. This is due to the problem of recovering residual common or joint costs that result from marginal cost pricing in utilities exhibiting both high levels of capital intensity and sub-additivity in the cost function. The means of recovering these residual common costs are discretionary, given the many possible economic solutions to the common cost problem. These include subsidies, multi-part tariffs, price discrimination, and volume discounts, as will be discussed in more detail in section 6.3 below.

The argument against implementing marginal cost pricing is surprisingly weak, and principally based on the contentions that marginal costs are too difficult to calculate and/or that utilities may not have a sufficiently accurate capital plan (see Fortin et. al., 2000). In reality, practical variations of marginal cost pricing methods are both readily available (see CWWA manual below) and simpler to use than cost of service methods, requiring only the classification of operating and capital costs as fixed or variable costs, compared to the complex process of cost and rate method allocation to user classes required under the fully allocated cost methods (e.g. base-extra capacity), as is also demonstrated below.

In terms of capital planning, water utilities are one of the most capital intensive of all utilities, providing basic services with regard to human and environmental health. They also have relatively predictable long-term water demands and costs. It is difficult to conceive of a well run water utility that should not have an adequate future capital plan, as is universally recommended in rate setting and water utility operating manuals. It is interesting to note that in England, each of 10 privatized water utilities is required to present a 20 year capital plan to the regulatory agency each five years in order to determine rates.

⁵⁴ Most industries in Ontario discharge their water-borne wastes to municipal sewers, as shown in Chapter Five

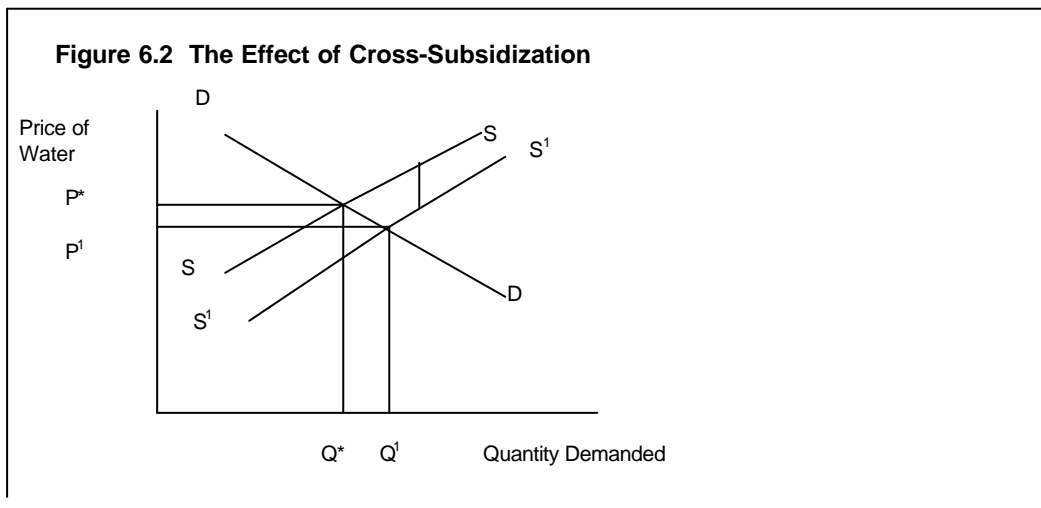
6.2.2 Full Cost Recovery and Subsidies

The issue of using subsidies to support water utilities is important, because unwarranted or unrecognized subsidies may interfere or even make impossible the achievement of efficient pricing practices. This section address two issues: (a) a description of current practices that create warranted cross subsidies, and (b) some guidelines that can be used for subsidization of water utility costs, which do not compromise the principal of full cost recovery, as outlined above. The possible efficiency of subsidization within water utility finance stems from the public goods nature of water utilities, the externalities associated with water pollution, and the problem of common residual costs.

a. Cross Subsidization of Water Services

Cross-subsidies in water systems occur when one water user's demand is financed by revenues derived from other users. Examples are common, and several have already been mentioned. Possibly, the most common case is that of flat rate water pricing, whereby all users in a given category of use (e.g., residential users) pay the same flat fee for unlimited access to public water servicing. In this case, the cross-subsidization is flowing from small-volume to large-volume users, and, in light of the discussion of the relationship between water demand and income, from lower to higher income groups. This is somewhat unexpected in a society that advocates progressive levels of taxation as an equity-inducing measure. Another example involves the use of promotional water rates for industrial water users. This involves a cross-subsidy from residential to industrial water users.

Theoretically, the mechanics of cross-subsidization is straightforward to illustrate (Figure 6.2). DD and SS are assumed to be the demand and supply curves respectively without cross-subsidies to this hypothetical user. The equilibrium price of water occurs at P^* , with equilibrium quantity at Q^* . The cross-subsidy has the effect of depressing the supply curve faced by the user to S^1S^1 , moving the new price and quantity to P^1 and Q^1 respectively. Clearly, the cross-subsidy both lowers the price of water to this user and raises the quantity demanded, which comprise movements away from the theoretical ideal of marginal cost pricing (represented by the original equilibrium.)



From the economic viewpoint, movements away from marginal cost pricing are generally undesirable, as dealt with in Chapter Two There are many cases, though, where cross-subsidization

may appear as justifiable. For example, consider the case of peaking characteristics shown by hourly, daily, and annual peaks patterns of the various municipal water uses. These peaking characteristics, combined with average demands, govern the overall size of water systems. As already demonstrated, peaking is generally more pronounced for the residential sector than for the commercial or industrial sectors. Under a common form of water pricing, the “base-extra capacity” method of the American Water Works Association (1983), this factor has been used specifically as the rationale for declining block rate structures. The argument is made that, because smaller water users (i.e., in the residential sector) are responsible for much of the peaking capacity in place, users in that sector should pay higher rates per unit of resource used than larger water users, who are not as responsible for the peaks. In other words, if peaking requirements were lower, water systems could be smaller.

This argument breaks down, however, because even large water customers draw on the supply during peak use periods, thereby contributing to peak demands. Simultaneously, by affording larger users cheaper per unit rates, either within the rate structure or under bulk water contracts, cross-subsidies occur from the small to the large water users. A simpler and more efficient pricing structure that would charge all users the same price per unit of usage, with this price based on the marginal cost of water supply to the system as a whole. This reflects views stated by Hirschleiffer et al. (1960), as well as by McNeill and Tate (1990), and in the Canadian Water and Wastewater Association (CWWA) rate manual (CWWA, 1993). It also stands in contrast to the pricing methodology suggested by Fortin et al. (2001), which appears to support average cost pricing. The idea that water price should be constant across user groups is not unlike pricing practices for most other goods and services in an economy.

There are other forms of cross-subsidies in a municipal water system, which are not easy to deal with. For example, geography can play a significant role in the cost of servicing. Housing units at higher elevations, for example, impose significant extra pumping costs, compared to those on flat ground. Also, serviced units at the extremities of systems entail higher costs. On the waste collection and treatment side of the water use cycle, high water users, or emitters of extra strength waste impose higher costs on treatment systems than average residential users. Theoretically, each of these factors gives rise to cross-subsidies, and could be allowed for in establishing water rates. However, these types of cross-subsidies are probably minor compared to those resulting from some water rate structures (e.g., flat rate pricing structures or DBRs), and, moreover, could be difficult to justify in equity terms.

b. Justifiable Subsidies in the Public Goods Context

As already demonstrated in Chapter Two, the public goods aspect of water utilities derives from the possible lack of private interest in providing adequate water services, and the common perception that potable running water is perceived by society as more of a right than a market good, as is also the case with in-stream water quality. Both the public goods nature of water utilities, and the existence of common costs that cannot be unambiguously allocated on efficiency grounds, appear to allow considerable scope for equity-based decisions in progressively allocating the common costs of water utilities to better-off members of society. This is consistent with social values as expressed in other progressive means of income distribution, such as income taxes or GST rebates, but contrasts with current rate setting methods, as embodied in the AWWA manual. The declining block rates advocated by the AWWA, as well as the common use of flat rates, are regressive, resulting in low income, low volume users paying proportionally more of the common residual costs for their water, because low income consumers typically have less access to high water use assets, for example, swimming pools. The AWWA methods also result in industry paying proportionally less than residential consumers. In contrast, assuming that progressive cross-subsidies are limited to common residual costs, there should be no implications relative to efficiency.

In the case of small communities, where scale efficiencies may be relatively small, cross-subsidies may be required to establish water system adequacy. Cross-subsidization, in this case, would be from the general taxpayer to the small communities. These cross-subsidies ideally would be one-time only events, and once adequate systems have been established, efficiency and full cost recovery principles should be established to assure on-going financial viability to the greatest degree possible. Another possibility, in appropriate geographic conditions (e.g., a relatively dense concentration of smaller communities), would be the establishment of regional water servicing. Studies by Tate (1972a, 1972b) showed that regional waste treatment systems would be cheaper to establish and operate than individual small community systems in the Yamaska and St. Francois Basins in Quebec.

6.3 Economic Theory Related to Water Utility Pricing

This section will review economic theory central to the issue of rate setting in utilities characterized by capital intensity and sub-additivity in the cost function. For simplicity of exposition, we assume a single utility providing one service (e.g. potable water), and thus a utility characterized by increasing returns to scale. The issues addressed will include: pricing methods for full cost recovery; dynamic pricing and optimal capacity expansion timing; and the definition of long run marginal cost.

6.3.1 Static Economic Theory and Full Cost Recovery

Static economic theory related to full cost recovery stems from the observation that in industries characterized by increasing returns to scale (or sub-additivity), marginal cost pricing will generate revenue shortfalls, as outlined in Section 2.2. The standard economic methods developed to deal with this problem can be broadly divided into two areas: first-best solutions, which rely on varying forms of lump sum transfers to achieve full cost recovery; and second-best (or quasi-optimal) solutions, which employ various forms of price discrimination. More recent work also discusses efficient public pricing in the case of utilities with decreasing returns to scale (e.g., as in the case where changes in water source arise).

Three charges, or sets of charges - volumetric, connection, and access-will be referred to throughout the chapter. These types of charges occur frequently in network industries.⁵⁵ Volumetric charges recover the marginal cost per unit of potable water or effluent. Connection charges recover the marginal costs⁵⁶ associated with connecting to the distribution network. Access charges recover the residual or common costs associated with sub-additivity in production that are not recovered through the other two charges.

The discussion below follows a rough chronological order. The initial work in the field dealt with developing first-best, or optimizing, solutions. This was followed, starting in the mid-1970s, with efforts to develop second-best solutions, normally involving price discrimination. The final part of the section discusses efficiency and equity considerations related to increasing block rates.

a. First-Best Solutions

⁵⁵ The concept of sub-additivity was defined in Chapter Two as, "... a situation where a single firm can provide a set of services at lower costs to all users than multiple firms providing separate services."

⁵⁶ An example of connection charges is development charges recovering the capital costs of connecting to the local distribution network.

The older first-best solutions are characterized by an emphasis on some form of non-distortionary lump-sum transfer of consumers' surplus to the utility. The challenge to managers is to recover the revenue shortfalls inherent in utilities with increasing returns to scale as efficiently as possible.

i) Taxation and Subsidization: One of the oldest solutions to the problem of full cost recovery these types of industries was advanced by Hotelling (1938), based on Dupuit's (1848) pioneering work justifying value in use, or consumers' surplus, as the correct criterion for determining the production of public works.

Given that optimal welfare is maximized in any one sector at prices determined by the equilibrium of supply (as reflected by the marginal cost curve) and demand, the solution advanced by Hotelling was to set prices equal to marginal cost. Revenue shortfalls would be recovered through non-distortionary taxes from general government revenues. Hotelling suggested that these non-distortionary taxes include inheritance, income, and economic rent (the latter arising principally, but not exclusively, from land ownership).

Hotelling postulated that subsidization financed from the suggested taxes would form an efficient redistribution of income, or an application of the second welfare theorem. He also suggested that the taxation and subsidization solution would have no income redistribution implications. Hotelling justified the neutrality of the income redistribution by postulating that, because public works are ubiquitous throughout a modern economy, taxes generated from one group of individuals to subsidize public works for a different group would be roughly equivalent to taxes imposed on the second group to subsidize works for the first group.

ii) The Coase Two Part Tariff: Hotelling's suggested solution initiated a debate, led by Coase (1946), who criticized the use of general taxation methods to subsidize public goods. The three basic criticisms advanced by Coase were: a) the policy would represent an inadvertent redistribution of income; b) income taxes were not lump-sum taxes but were distortionary in their own right; and c) pricing solely at marginal cost would distort the allocation of resources in favour of the subsidized utility.

The criticism of income redistribution posited that it would be unlikely for the allocation of subsidies to decreasing cost industries to result in a rough trade-off in income redistribution among individuals. The example Coase used was the case of public utilities with a predominately urban clientele being subsidized by the rural population. Coase also argued that the income tax itself is distortionary because of its progressive nature, and because the tax constituted a type of excise tax on effort which is sub-optimal.

The criticism on the basis of misallocation of resources was basically a benefit-cost criticism, whereby the existence of a utility is not a foregone conclusion, and utilities may either start up or shut down (Laffont and Tirole, 1993). A common example of this, in terms of water utilities, would be the decision by small communities to continue to self-supply (i.e. wells) or move to public utility supply. In order for the government to know that the optimal allocation of resources occurs through public utility supply, potential consumers must be charged all of the costs incurred by the utility, and then allowed to make the decision to connect, or not, to the utility. If the full costs were not attributed to the consumers potentially served by the utility, as in Hotelling's proposal, consumers might choose to obtain supplies through a public utility, even though a more efficient solution from a social perspective would be continuing self-supply. This occurs because the subsidy distorts the costs borne by the consumer.

Coase suggested using a two-part tariff in place of the non-distortionary tax concept advanced by Hotelling. Basically, a two-part tariff is a pricing system whereby each consumer (in this case of water services) would be charged a single price per unit of output (e.g., per cubic meter of water service) purchased, but, in addition would pay a lump-sum, or fixed charge, for the opportunity to be

able to buy any unit of the service. The price per unit, or volumetric charge would, of course be based on the marginal cost of providing the service.

iii) Criticism of Coase's Pricing System: Coase's work initiated an additional debate, reflecting the complexity of pricing public goods. Vickery (1948) suggested that Coase had misinterpreted the problem of recovering residual common costs by confusing one type of marginal cost, the marginal cost of the distribution network, with residual common cost. Coase had advanced a hypothetical situation in which a good was produced by a central production facility with a marginal cost of production of zero, and then conveyed to each consumer through a distribution system consisting of one carriage per consumer. Coase had criticized the Hotelling proposal on the grounds that the consumer would pay the marginal cost of production, but not the cost of the carriage (which Coase terms a fixed cost), and thus resources would be allocated in greater quantities than optimal to the given utility.

Vickery suggested that the correct interpretation of the problem posed by Coase was simply that two goods were being provided to the consumer, one the good itself, and the second a means of conveying the product to the consumer, in other words a distribution network. In the Coase example, there were no increasing returns to scale. The consumer would pay the marginal cost of production (volumetric charge), as well as the connection charge, or marginal cost of the distribution network, and optimal production results. The two goods (production and distribution) are complementary goods with some degree of cross-price elasticity in the long run, where a decision on connecting or disconnecting to a public utility can be made.

It was also argued that Coase's analysis also misinterpreted the problem of increasing returns to scale in production, where there is a residual common cost (i.e., distinct from the connection and volumetric charges) that must be allocated and recovered in order to achieve full cost recovery. In this context, an interpretation of the Coase two-part tariff would be that the residual common cost is recovered through an access charge imposed directly on the population served by the utility.

The access charge can be seen as a form of direct lump sum or non-distortionary transfer to the utility, if the cross price elasticity between the volumetric, connection and access charge, is zero, or alternately if water demand related to the access charge is perfectly inelastic (McNeil, 1989). Thus the problem for utilities in implementing the access charge is to ensure that an access charge on any given connection will not result in a particular user refusing to connect, or disconnecting, from the municipal system, thereby generating cross-price elasticities between the access charge, and the volumetric and connection charges.

iv) Conditions for Connecting to the Public Water System: Wiseman (1959) pointed out that the decision to connect to a public utility is typically a discrete technology choice due to the substantial fixed costs involved. In terms of water supply this discrete choice is generally between self-supply, and connecting to a local public utility. Thus an access charge limit-pricing rule for any given connection can be formulated as;

$$(6.1) \quad A_j < TCS_j - VQ_j - C_j$$

where: A = annual access charge;
TCS = annual costs of self-supply for the j^{th} customer;
V = annual volumetric charge;
Q = expected annual water demand for the j^{th} customer;
C = annual connection charge for the j^{th} customer; and
j = the number of potential connections to the utility

Non-negativity constraints are assumed for A, C, V and TCS. This limit pricing constraint is essentially the same as that developed in Ng and Weiser (1974)⁵⁷.

The pricing limitation rule of equation 6.1 ensures that, for all potential connections, the annual aggregate marginal costs of being served by a given public water utility are less than the costs of annual self supply. In the circumstance where an access price of zero cannot generate $TCS < VQ + C$ for a given connection the social optimum is generated by potential connections continuing to self supply rather connecting to the utility.

In the case of water supply the pricing limitation rule indicates that two customer classes could potentially pay a lower proportion of residual common costs (or a lower access charge) for centralized water servicing: large industrial consumers or industrial parks, and outlying rural households or communities at the edge of a growing urban area. This occurs because both of these customer classes have viable self-supply alternatives. The two consumer classes that may pay a higher proportion of residual common costs are urban residential and commercial users. It should be noted that the access-limit pricing rule is a form of Ramsey pricing, as described in the next section.

b. Second-Best or Quasi-Optimal Solutions

More recent analyses have given less emphasis on optimal solutions to the cost recovery problem arising from the nature of public goods. They have tended, instead, to concentrate on developing second-best or quasi-optimal solutions to the cost recovery problem. Two points characterize these solutions: the use of price discrimination as the means of recovering system costs; and a stronger emphasis on mathematical techniques. The two broad methods developed are Ramsey pricing, and "Pareto Superior Non-linear Outlay Schedules" or volume discounts (Willing, 1978).

i) Ramsey Pricing: The technique of Ramsey taxation was first developed by Ramsey (1927). Baumol and Bradford (1970) provided a further analysis of this solution. Ramsey prices attempt to maximize welfare subject to full-cost recovery, through the use of third degree price discrimination⁵⁸ based on the differing demand elasticities of the various user classes⁵⁹. Where water supply is considered to consist of only one good (i.e. potable water) and there are no cross price elasticities of demand between user classes (as is the case for water supply) the pricing rule for each user class can be represented as:

$$(6.2) \quad R = V_i - (MC_i / V_i \cdot E_i)$$

where: R = the constant mark-up needed to achieve full cost recovery (this is termed the Ramsey number);
V = the volumetric charge for the service;
MC = the service's marginal cost;
E = price elasticity of demand; and
i = user classes indexed from 1... .i; subject to $0 < R < 1$.

⁵⁷ The mathematical proof of the first best optimality of this pricing rule can be found in (Crew and Kleindorfer, 1979).

⁵⁸ "Third degree price discrimination occurs when consumers are charged different prices, but each consumer faces a constant price for all units of output purchased" (Varian, 1992)

⁵⁹ An alternative form of Ramsey pricing involves second-degree price discrimination. By classifying each segment of incremental quantity supplied as a separate good, and assuming independent demands for each good, Ramsey prices can be calculated for each separate market. This will lead to declining block rates in markets with increasingly elastic demands, or alternatively increasing block rates in markets with increasingly inelastic demands (Brown and Sibley, 1986). In equation 5.1 above this would mean that incremental quantity consumers, rather than user classes, would be indexed from 1...i.

When Ramsey numbers are 1 the outcome is identical to that of a price discriminating monopolist, while a Ramsey number of 0 indicates perfect competition. Thus, Ramsey numbers between 0 and 1 indicate the use of some degree of monopoly power in order to ensure full cost recovery (Scherer, 1990).

The inverse elasticity rule demonstrated above is the mathematical representation of the intuition that the maximization of welfare subject to full cost recovery results from charging consumers with inelastic demands a higher price relative to consumers with elastic demands, thereby resulting in the smallest possible reduction in optimal production.

The application of Ramsey pricing becomes somewhat more complicated where water utilities are considered to provide two or more goods or services (i.e. production and distribution, water and sewage), and where these goods have some degree of cross-price elasticity. In the two good cases, using a utility that provides solely potable water production and distribution; the pricing formula becomes:

$$(6.3) \quad R = \frac{V_i - MCP_i}{V_i} \cdot (E_{vi} - E_{cvi}) = \frac{C_i - MCD_i}{C_i} \cdot (E_{ci} - E_{vci})$$

- where: R = the constant mark-up needed to achieve full cost recovery (or the Ramsey number);
V = volumetric charge;
MCP = marginal cost of production;
MCD = marginal cost of distribution;
C = connection charge;
E = price elasticity of demand;
i = user classes indexed from 1... i, subject to 0 < R < 1.

The general elasticity terms combining both the own price elasticity, and the cross price elasticities can seen as a form of net price elasticity (Train, 1994).

The use of Ramsey pricing means that a price mark-up above marginal cost for all charges (e.g. connection and volumetric charge) will incorporate the residual common costs recovered in the Coase solution through the use of the access charge. Thus the use of Ramsey pricing means that no access charge need be instituted by the water utility.

The use of Ramsey pricing may lead to problems for rate makers in terms of common perceptions of equity (Train, 1994). This occurs because, in general, consumers with inelastic demands for a good have a lack of substitutes, while consumers with a greater elasticity of demand have a greater choice of substitutes. To use the example of urban mass transport, under Ramsey pricing, low-income consumers without cars would be charged higher rates to ride the bus, while middle or upper income consumers with cars would be charged lower rates. In the case of water supply, industrial consumers or high income suburban residential users with higher demand elasticities would be charged lower rates, while low-income urban residential consumers would be and are charged higher rates.

An interesting observation related to Ramsey pricing is the similarity of this form of pricing to the declining block rates developed in the AWWA manual. This occurs because residential consumers have lower demand elasticities relative to industrial consumers, as discussed in Chapter 4. An interesting econometric study by (Kim, 1995) referred to in Chapter 5 indicated that the price mark-up over marginal cost for industry in a sample of U.S. water utilities was relatively consistent with the Ramsey numbers required for full cost recovery.

ii) Pareto Superior Non-Linear Outlay Schedules: Perceived problems of equity inherent in Ramsey pricing arise because welfare improvements arising from this type of pricing, when compared to average cost pricing, may generate both winners and losers in society (Willig, 1978). In terms of water supply, the potential losers are low-income residential users, while the potential winners include large industrial concerns. While, theoretically, the winners could compensate the losers, and still retain some surplus, without a socially mandated compensating mechanism outside of the utility, it is doubtful this would occur.

A Pareto superior⁶⁰ pricing schedule that can ensure full cost recovery will avoid this equity problem by guaranteeing that the losers are at least no worse off than they were under uniform average cost pricing, and are potentially better off depending the pricing policy imposed on the utility. The Pareto improving solution advanced by Willig (1978) addressed these concerns through the development of a declining block rate structure based on volume discounts, a form of second-degree price discrimination⁶¹.

In his paper Willig proved that, under the standard assumptions governing firm and consumer behaviour:

- for any uniform price greater than marginal cost, there exists a declining block rate structure which Pareto dominates the uniform price, and which generates greater utility profit; and
- a Pareto efficient block rate schedule must offer the largest consumer a price equal to marginal cost.

These conclusions are subject to the conditions that the product cannot be readily traded amongst consumers, and that the vendor must be able to monitor individual purchases⁶².

The simple graphic logic behind the proofs advanced by Willig is demonstrated in Figure 6.3, in the context of a utility where all costs must be recovered solely through volumetric charges. At P_1 the uniform, or average cost price, the utility produces at Q_1 , and fully recovers its costs. By adding a decreasing block rate priced at P_2 total consumption increases by the square BCQ_1Q_2 , consumers' surplus increases by area ABC , and utility profits increase by area $BCDE$.

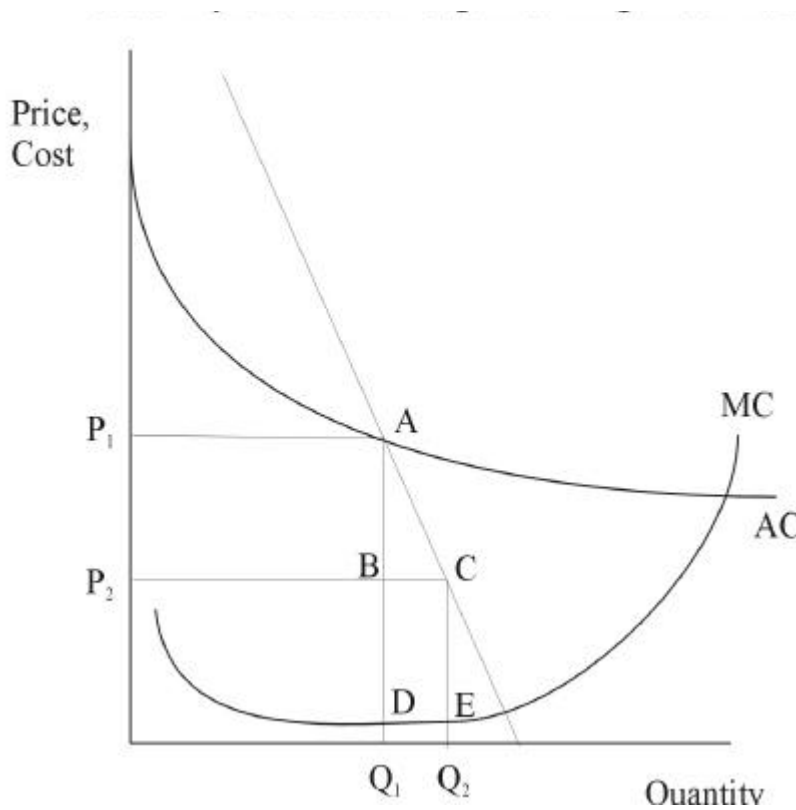
If the utility is constrained to break even, moving to the declining block rate while maintaining the original price (AC) increases profits ($BCDE$), and thus the price in the first block must decline to below AC , and a Pareto improvement occurs for all consumers, while the utility continues to recover its costs. The same logic is involved in adding new blocks until the final block, or rate charged the marginal consumption of the largest consumer, must equal marginal cost.

⁶⁰ Recall that Pareto optimality results when resources are distributed in such a manner that no-one can be made better off (through re-distribution) without making someone else worse off.

⁶¹ "Second degree price discrimination occurs when prices differ depending on the number of units bought, but not across consumers" (Varian, 1992).

⁶² An additional condition advanced by Ordover and Panzar (1979) is that users demand for final goods must be independent.

Figure 6.3 Pareto Improvements through Declining Block Rates



Willig also advanced the proposition that while the Coase result is welfare maximizing, given the possibility of lump sum transfers, the implementation of a two (or three) part tariff may result in the same equity problems referred to under Ramsey-Boiteux pricing. These equity, or Pareto optimal, concerns stem from a two or three part tariff potentially increasing the total costs paid by a given connection compared to average cost pricing.

iii) Increasing Block Rates: Increasing block rates differ from the previous rate setting methods because they represent an efficient pricing solution to the problem of cost recovery in increasing returns to scale utilities. They solve the problem of excess profits as generated by diseconomies of scale generated by scarce water resources, in particular, droughts (Hall and Hanneman, 1996). They were developed in the context of Southern California climate and geography. Increasing block rates act to restrict demand by charging high volume water users proportionally more than low-income users, thus choking off demand quite severely, particularly where high volume users also exhibit higher price elasticities. The structure of the tariff is to charge high volume users the marginal costs of the water, and set the lower volume rates so that full cost recovery ensues, essentially the reverse of the Willig's argument, as cited above.

In the case of water utilities, this structure also generates an equity transfer from high volume water users, typically better-off members of society (e.g. swimming pool owners) to lower volume users.

The equity argument was advanced by (Feldstein, 1972) as a justification for possible increasing block rates in a comment on the regressive impacts of possible flat connections charges in the two-part tariff. Increasing block rates are also often advocated on the grounds of water conservation.

6.3.3 Peak Load Pricing and Dynamic Optimal Capacity Timing Decisions

The use of peak load pricing stems from the observation that demand is not evenly distributed over time, and that capacity must be sized to meet maximum or peak demands. The peak load problem is to determine the time differentiated pricing schedule that results in optimal capacity. Typically analysis of this issue is conducted on two levels, continuous capacity, and discontinuous capacity, a differentiation used in this section. The first analyst to generalize the theory of peak load pricing to the standard demand and supply analysis was Williamson (1966), although Boiteux (1949), Steiner (1956), and many others preceded him in recognizing its importance.

Additional complications related to the analysis of peak loads and optimal capacity, such as the concept of shifting peaks (Pressman, 1970), diverse production technologies (Crew and Kleindorfer, 1976) and possible storage (Nguyen, 1976) will not be addressed. In the case of water supply, the shifting peak problem can be safely ignored because both metering constraints, and the standard practice of installing storage to meet the daily peaks generate a single seasonal peak period that can be used for peak load pricing. Since the seasonal peak is determined primarily by seasonal activities (e.g. lawn watering, swimming pools), it will not shift with changes in price, though the level of the peak will be determined by the peak load price.

Peak load pricing and storage will not be addressed for the same reason. Current practice with respect to storage in the local distribution network leaves only a seasonal peak, as the storage in the distribution network is designed to meet the peaks in daily demand, and provide a reserve for fire protection. The main trunk component of the system is designed to meet this seasonal demand peak (maximum daily demand). Although in arid regions storage may also be built into the main trunk component, this storage is designed to meet water shortages, rather than to even out peak demands, and the storage (e.g. reservoir) will also be sized to meet maximum daily demand. The problem of diverse production technologies and optimal plant mixes in supply can be ignored due to the homogeneity of production technologies in water supply, and due to the economies in production that characterize the industry.

A major difference from the standard treatment of peak load pricing will be the explicit use of an increasing returns to scale long run production function (or to cope with discontinuous capacity, a long run production function exhibiting sub-additivity), as is consistent with natural monopoly. One curiosity with respect to the standard treatment of peak load pricing⁶³ is the use of a long-run constant returns to scale production function, and the transfer of conclusions derived from this analysis to public utilities. Though this assumption does simplify the use of mathematical techniques such as static optimization, dynamic programming, and optimal control theory, it tends to generate results differing substantially from an analysis under long run sub-additivity.

In light of the mathematical difficulties encountered in dealing with increasing returns to scale in a dynamic context, the analysis in this chapter will be confined to graphics.

The properties used throughout this chapter in terms of the shape of the short (SRMC) and long run marginal cost (LRMC) curves are taken from Varian (1986), and the econometric studies surveyed in chapter 5 of this report. These properties are:

⁶³ An exception can be found in the static analysis by Mohring (1970), with results similar to those presented here in sections 6.2 and 6.3.

- $MC = AC$, at the first unit of production, and at the point of minimum average cost;
- the assumed continuity of LRMC, $LRMC=SRMC$ at optimal plant size for a given level of demand;
- limited substitutability between capital and short run variable factors generate a cost inelastic SRMC at or near designed plant capacity (i.e. a fixed capacity plant [Rees, 1984]).

These assumptions serve to generate a short run marginal cost curve which is decreasing at low levels of capacity utilization, relatively smooth in the transition from decreasing to increasing marginal costs, and then increasing in the portion of the short run marginal cost curve near designed capacity of the production plant. The econometric studies cited earlier indicate that the general case for water utilities is that of decreasing short run average cost.

- the marginal physical product of capital is increasing in quantity supplied.

This characteristic is derived directly from the econometric studies surveyed in Chapter Five, which indicated that the 2/3 engineering rule will generally hold in potable water production.

a. Assumed Continuous Capacity

In this section, a continuous LRMC is assumed to exist, and the static graphic analysis will demonstrate the relationship between optimal capacity, the pricing schedule at peak and off-peak demand, and the relationship between the demand for capacity and the marginal cost of capacity.

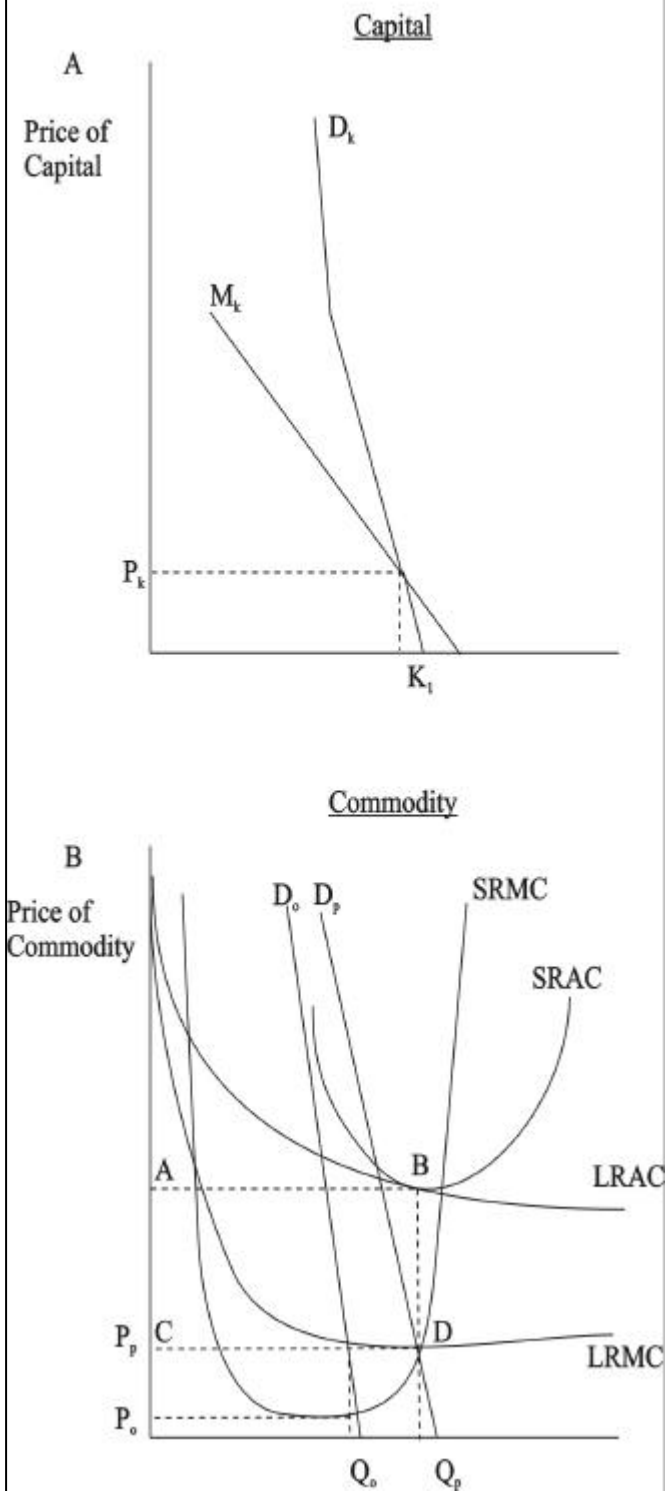
Corresponding to the standard static analysis (demonstrated mathematically in Appendix 6.A1), the derivation of the optimal schedule of peak and off-peak prices stems from capacity as a joint cost shared by different users in different time periods. As the use of capacity by consumers in any one period does not impinge on the use of capacity in other periods, consumers' surplus is additive, and can be seen as the sum of the two consumers surpluses relative to capacity in the two time periods. Thus a demand for capacity curve can be constructed as the summation of the two demand curves for water above the short run marginal cost of production.

In Figure 6.4, the demand for capacity is represented by D_c , while the two commodity demand curves are represented by demand curves D_o and D_p for off-season demand and peak demand respectively. Equating the declining marginal cost of capacity to the demand for capacity generates an optimal capacity level of K_i , resulting in a peak price of P_p , an off-peak price of P_o , and access charges whose revenues must be equivalent to the area (ABCD).

As stated in the previous section, an access charge⁶⁴ must be imposed in addition to the peak load prices recovering joint marginal costs. The access charges recovers the common costs generated by utilities operating in the increasing returns to scale portion of LRAC. The imposition

⁶⁴ A second best solution to recovering the residual common costs (ABCD in Diagram 6.1) in an industry that also exhibits joint cost characteristics is Ramsey-Boiteux pricing. This pricing method uses third degree price discrimination involving demand elasticities to mark up prices above the peak load prices recovering joint marginal costs. A demonstration of this pricing method can be found in Laffont and Tirole (1993).

Figure 6.4 Peak Load Pricing with Continuous Capacity



of the access charges is a departure from the standard treatment where the peak and off-peak volumetric prices generate sufficient revenues for full-cost recovery

Figure 6.4 also demonstrates that, under increasing returns to scale, LRMC the peak load users must pay both the short run operating costs, and the capital or capacity costs of the utility (LRMC), while off-peak users will pay only the SRMC. Again, this contrasts to the standard treatment using constant LRMC, where depending on the relative position of the peak, and off-peak demands the capacity or capacity costs may be shared between the two period demands. This result arises because optimal capacity occurs solely at $LRMC=SRMC$, and the curvature of the SRMC, and LRMC curves guarantee that at optimal capacity only one demand curve can intersect both SRMC and LRMC.

This section has assumed demand lies in the continuous increasing returns to scale section of LRAC, and generated a graphic, static solution to the peak load-pricing problem. However the assumption of a continuous LRMC and a static solution is inappropriate in the case of an industry facing dynamic growth in demand, and possessing either economies of scale, or economies related to capital. The next section will discuss the problem of peak load pricing in the context of optimal discontinuous capacity.

b. Discontinuous Capacity

Where demand is exogenously increasing over time, and an industry exhibits economies of scale, or economies related to capital, the optimal schedule of capacity expansion becomes discrete, and the LRMC curve disappears, being replaced by a succession of SRMC curves corresponding to the discrete levels of optimal capacity. Manne (1961) was the first author to derive a mathematical solution to this dynamic problem, using the mathematically necessary assumptions of a constant exogenous rate of growth in perfectly inelastic demand, no depreciation of capital, constant long run returns to scale, and a Cobb-Douglas specification of the production function.

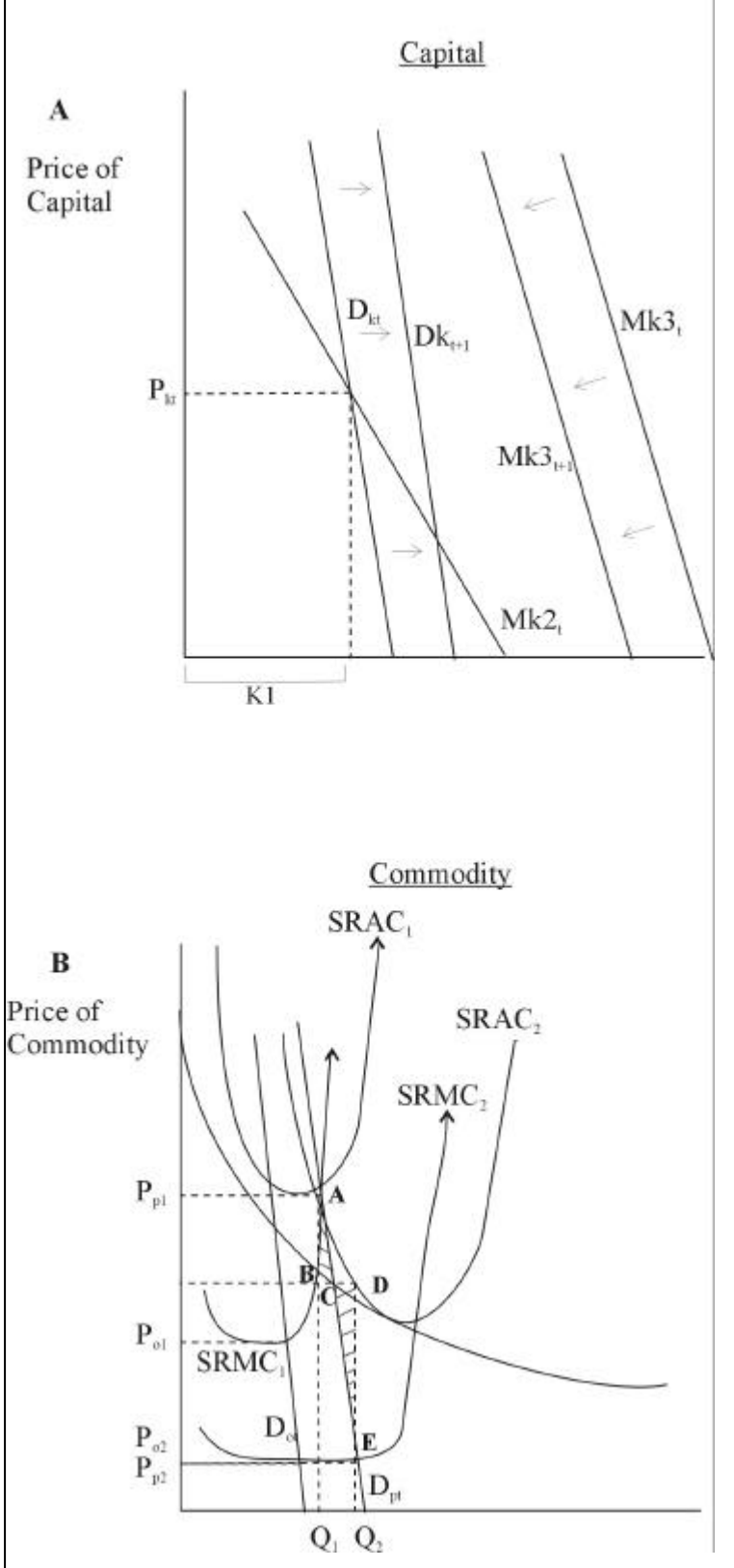
The mathematical solution to this dynamic programming problem is shown in Appendix 6.A2.

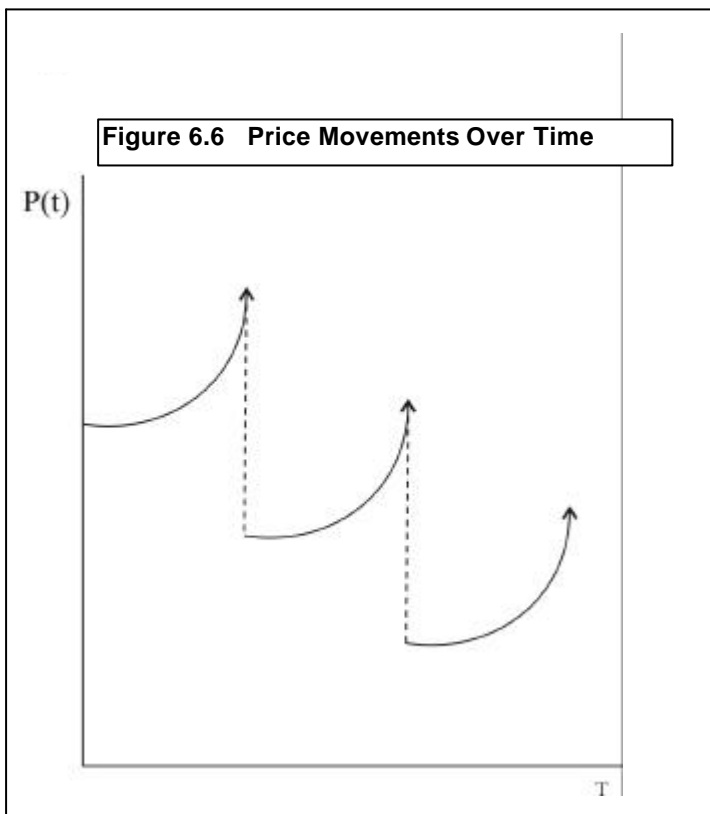
The graphic analysis in this section (Figure 6.5) relaxes these assumptions, notably those related to demand, and depreciation. The graphic analysis will attempt to portray the dynamic process for determining optimal capacity, and peak load pricing, through comparative movements in time of the peak and off-peak demands, and the marginal cost of capacity.

The assumptions used are that marginal capital cost, demand for capacity, and peak and off-peak demands are linear. Peak and off-peak demands are assumed to be increasing over time. The rate of depreciation is assumed to be constant over time. The marginal cost of capital at any point in time is assumed to be a function of the existing stock of capital, while time is assumed to be discrete, and represent the end of the short run time period needed to build a capital facility. It is also assumed that the potential design capacity of any given plant is continuous, rather than being indivisible, or restricted to fixed capacity increments.

In Figure 6.5b, the short run average cost curves ($SRAC_1$, $SRAC_2$) for two plants represent optimal discrete levels of capacity. At D_{pt} , the utility is at the discontinuity where the changeover from plant 1 ($SRAC_1$) to plant 2 ($SRAC_2$) takes place. At time t , the net increase in consumers' surplus (CS) from building plant 2, the triangle (ABC), is exactly equal to the loss in producers' surplus (PS), the triangle (CDE). This optimal capacity timing decision was demonstrated mathematically by Rees (1984), based on a graphic analysis by Williamson (1966), using the standard assumption of long run constant returns to scale, as well as the assumption of fixed capacity increments, or indivisible capital.

Figure 6.5 Peak Load Pricing and Discontinuous Capacity





Corresponding to this capacity timing decision, as demonstrated in figure 6.5a, at time t the demand for capital D_k intersects the declining marginal cost of capital curve (Mk_2). Because plant 2 (i.e., as represented by $SRAC_2$) has just been built, at time t the new marginal capital cost function becomes Mk_3 , where the discontinuity arises, as the marginal cost of capital is a function of the existing capital stock, which has just increased by a discrete amount. The initial unit of Mk_3 represents the cost of adding one more unit of capacity, given that plant 2 has just been built, and has thus become a fixed rather than a marginal capacity cost.

.As all demands are increasing over time, at time $t+1$ $D_{k,t+1}$ moves horizontally out from the y-axis, while $Mk_{3,t+1}$ moves horizontally toward the y-axis as depreciation erodes the value of the existing capital stock (i.e., plant 2). The intersection of these two curves (D_k, M_k) at same time $t+n$ will then generate the next increment in capacity, plant 3, as would be represented by a new short run average cost curve, $SRAC_3$ (not shown).

Thus, the relationship between depreciation of the current capital stock, growth in demand, and an initial level of capacity will determine a succession of optimal discontinuous short run average and marginal costs both over time, and in relation to production. Figure 6.5b), demonstrates the discontinuous pricing schedule for peak (P_{p1} to P_{p2}) and off-peak demand (P_{o1} to P_{o2}) at time t , where both peak and off-peak prices correspond solely to SRMC. An access charge is also needed at AC_2 to ensure full cost recovery, and profits are being generated at AC_1 .

As demonstrated in 6.5 an interesting observation related to peak load pricing at SRMC is the potential for off-peak prices to be higher than those in the peak, where off-peak demand lies in the range of decreasing SRMC. This is possible in the case of water utilities given that SRAC is decreasing with the rate of capacity utilization.

Figure 6.6 demonstrates the stylized movement in the peak prices over time, where the potentially rather severe price movements corresponding to pricing at SRMC in the context of optimal discontinuous capacity can be seen as an example of the "responsive pricing" advocated principally by Vickery (1955, 1971)⁶⁵. These price movements will be more extreme the more inelastic demand, and the larger the optimal discrete capacity increments over time, with both conditions characteristic of water utilities.

6.3.4 Rehabilitating Long Run Marginal Cost Given Discontinuous Capacity

In the previous section the optimal capacity expansion schedule under the cost conditions normally associated with public utilities yielded a discontinuous succession of short-run marginal cost curves as a replacement for continuous long-run marginal cost curve. The optimal volumetric price under these conditions equates demand solely to short run marginal costs. These results generate substantial price instability over time, particularly at the point of the discontinuity, and suggest that in most, if not all, time periods an additional access charge will be needed in order to achieve full-cost recovery. The analysis relies on the standard economic definition of the total cost of capital, where the cost of capital at any point in time is the rate of return on existing capital, plus the rate of depreciation, given no taxation of capital. Both these results, and the standard definition of capital costs which underlie them, have been criticized as inadequately incorporating time within the marginal capital cost function (Turvey, 1969), as well as generating socially unacceptable rate instability over time (Mann 1981).

This section will review a practical means of calculating LRMC, given discontinuous capacity, based on (Turvey, 1969). The definition of long run marginal cost as the sum of short run marginal cost (or marginal operating cost) and marginal capital cost as in equation 6.4:

$$(6.4) \quad \text{LRMC} = \text{SRMC} + \text{MCC}$$

where: MCC = marginal capital cost.

Other characteristics defining the long run are:

- the interpretation of the long run as the time period corresponding to the next increment in capacity;
- the use of a discrete unit of time (current year) for the analysis;
- the assumption that inelastic demand is increasing over time; and
- the assumption that all capacity costs are perfectly variable with respect to changes in demand.

The assumption of discrete time generates a definition of SRMC equivalent to the change in total operating costs divided by the change in expected demand, for a given year, given fixed capital costs. This is represented in equation 6.5 below;

$$(6.5) \quad \text{SRMC}_t = \Delta \text{OC}_t / \Delta Q_t$$

where OC = operating costs;
Q = expected demand; and

⁶⁵ An empirical analysis exploring the potential welfare implications of moving from dynamic SRMC pricing to a stable price based on a weighted average of marginal costs can be found in (Swallow and Mann, 1988). The results indicate a 1.5 ~ welfare loss from moving to stable prices.

T = time.

a. Turvey Marginal Cost

The concept of Turvey marginal cost (Turvey, 1969 1976) is based on the axiom that, given some growth in demand, additional capacity increments cannot be totally avoided, but can be postponed (advanced) with reductions (increases) in annual demand. Therefore the marginal capital cost becomes the change in the present worth of the next increment in capacity divided by the change in annual demand necessary to postpone (or advance) the building of that capacity increment. In order to reduce this to the discrete time unit of the current year, Turvey assumed that the appropriate annual increment in demand to be used would be equal to the increment expected in the current year. Avoiding this increase in demand can be assumed to generate a postponement of the next increment in capacity by one year. Thus Turvey long run marginal cost can be represented as;

$$(6.6) \quad MCC_t = \frac{pr_{t_0} - pr_{t_0+1}}{\Delta Q_t}$$

Where p is the standard discount factor;

$$(6.7) \quad p = \frac{1}{(1+d)^m}$$

And m represents the time period from the next capacity increment to t, while d is the discount rate.

The variable r refers to the standard amortization factor;

$$(6.8) \quad r = \frac{i(1+i)^{t+n}}{(1+i)^{t+n} - 1}$$

Where i = interest rate, and n = life expectancy of the capital equipment.

b. An Extension to Turvey Cost

While the Turvey method represents a major development in both conceptualizing, and estimating long run marginal cost, it can be criticized for the unduly limiting assumptions that the long run consists solely of the next increment in capacity.

Thus an extension to the Turvey concept can be specified as:

$$(6.9) \quad MCC_t = \frac{\Delta \sum pr_t}{\Delta Q_t}$$

This generalization allows for a more practical application to utilities, where a capital investment plan will typically comprise an uneven series of capital investments in different types of equipment, in different time periods.

An interesting observation stemming from the Turvey concept is the specification of an arbitrary change in hypothetical current demand, and the necessary assumption⁶⁶ of a relationship between the hypothetical change in demand, and the present value of total capital cost. The assumption

⁶⁶ This assumption is needed because actually predicting how future capital projects will shift due to small changes in hypothetical demand is difficult. It is difficult due to the presence of long run fixed costs, which may generate discontinuities where a small change in demand generates no change in capital timing.

used by Turvey is that a reduction in Q_t by the annual increment in demand generates a delay of one year in all capital projects.

Given that total capital cost (as defined in the numerator in equation 6.9 is some function of current demand, and the future rate of growth in demand (i.e. $TCC_t = f(Q_t, g)$), marginal capital cost can be either increasing, constant or decreasing, in relation to changes in current demand Q_t . An extension of the Turvey assumption indicates a decreasing relationship in Q_t , due to the use of the discount rate p . For example doubling the decrease in the expected increment in annual demand will generate a proportionally smaller change in marginal capital cost, as capital projects become postponed by two years into the future. This generates a series of marginal capital costs rather than a single number, depending on the specification of the change in Q_t .

An alternative, and simpler assumption, which resolves the problem of choosing an arbitrary hypothetical change in Q is the assumption of a constant relationship, or constant returns to hypothetical changes in current demand. This assumption has the convenient property that MCC now becomes;

$$(6.10) \quad MMC_t = \sum^t pr_t / Q_t$$

The constant returns to hypothetical changes in current demand assumption should be clearly differentiated from assuming constant returns to scale, or constant economies related to capital. The above stated assumption takes as given a particular optimal capacity expansion schedule that may reflect both increasing returns to scale, and/or capital economies. The constant returns refer to delaying or advancing those projects in time due to hypothetical changes in current demand, and not to the underlying relationship between production and cost.

The underlying relationship between production and cost will generate a movement in LRMC similar to that of SRMC in section 6.3.2, where LRMC will rise prior to a capacity expansion, and fall after that capacity is built, but with less extreme movements. This smoothing occurs when total capital cost becomes a discounted moving average of all future capital costs within the chosen long run planning horizon.

6.4 Practical Methods and Issues in Water Pricing ⁶⁷

The theory of rate setting outlined above cannot be applied directly to establish water rates, but rather establishes a foundation of principle. This foundation may then be taken by practitioners to establish rates that meet these principles to the extent possible, given the myriad other factors that have to be taken into account by any particular municipality.

The following discussion outlines two such methods – one a traditional approach used throughout North America, and the second a much newer and more experimental approach. Respectively, these are:

- the American Water Works Association (AWWA), “Rate Setting Manual, M1”
- the Canadian Water and Wastewater Association (CWWA), 1993, “ A New Approach to Rate Setting”

⁶⁷ This section provides a theoretical outline of two rate setting methods. It is not a detailed manual on these methods. The latter can be found by consulting the source documents.

Throughout this discussion, it is assumed that the municipality wishes to recover full costs, is interested in doing this efficiently, and is not cross-subsidizing other municipal services through its water charges.

6.4.1 AWWA Manual

The AWWA has been interested in establishing standards for rate setting since at least the early 1950s. Its "M1" manual is probably the best known and most widely used product of this effort. Although this manual has been criticized many times for its embedded inefficiencies and sometimes-arbitrary nature, it has nevertheless been widely adopted.

Two principal two-rate setting approaches are recommended in the AWWA manual: the base-extra capacity method (BEC); and the commodity-demand method. These methods are "cost of service" based methods based on four principal criteria; equity, rate stability, practicality and full cost recovery, where equity is defined as each connection paying the costs which it imposes, or has imposed, on the system.

These two rate-setting methods deviate from marginal cost pricing by three fundamental characteristics:

- Rates are set to recover the current operating and capital costs of the utility, without considering future capital needs of the utility.
- All current costs are fully allocated or assigned specifically to various service classes (e.g. residential, commercial, industrial).
- Current system costs related to water demand are recovered through volumetric rates based on average cost pricing.

The BEC method allocates historical costs to user classes (e.g. residential, industrial, commercial) based on a allocation of system costs to four categories of costs, referred to as rate method classes. In the BEC method, these rate method classes are the base, extra capacity, customer service and fire protection class. The volumetric charge recovers the base and extra-capacity costs, the fixed customer service charge recovers cost such as billing, while fire protection charges are recovered through property taxes. Table 6.1 provides an example of system costs allocated to the rate method classes.

Table 6.1 An Example of Cost Allocation using The Base-Extra Capacity Method						
Item	Total Cost \$ 000	Allocated to Base	Allocated to Extra Capacity		Customer Costs	Fire Protection
			Max. Day	Max Hour		
Source of Supply	800	800				
Pumping Plant						
- raw water	200	132	68			
- treated water	500	330	170			
Treatment Plant	1500	990	510			
Transmission	2000	880		1120		
Meters	300				300	
Fire Hydrants	150					150
General plant	350	201	48	72	19	10
Total plant value	5800	3333	796	1192	319	160

Base costs refer to those costs attributable to meeting average demands – that is, costs not associated with peaking – and include both O&M and capital expenses. Extra capacity costs are those associated with meeting demands over and above the average. Calculation of the split between these two types of costs occurs based on two ratios: average day to peak day (on a yearly basis); and average hour to peak hour usage. Both of these ratios can be measured from basic utility records. As an example, if the average: peak day ratio is 1:1.5, 66% of costs related to this peaking component are allocated to the base, 34% to extra capacity. Similarly, if the average: peak hour ratio is 1:2.5, 44% of costs related to this peaking component are allocated to the base, 56% to extra capacity.

Customer costs are associated with serving the customer, regardless of demand volumes. These include meter installation and reading, billing, management, etc. These costs are basically fixed on an annual basis. The AWWA tries to allocate fire protection costs solely to that function. Basically these costs include fire hydrants, and associated branch mains and valves. A considerable proportion of extra capacity may be used from time-to-time for fire protection. Accordingly, costs allocated to fire protection may be small.

The cost allocation to user classes is based on unit of services (US). An example of a US is volumetric water demand, the US used for allocating costs in the base and extra capacity rate classes. The US is used to calculate a set of unit costs of service (UCS), for example, base capacity cost per metric cube of base water demand. This UCS is then multiplied by the relevant service units assigned to each user class, for example, base water demand assigned to the residential user class, in order to calculate the rate method class cost allocated to the user class, for example, the proportion of base capacity cost allocated to the residential user class. These costs are then summed by user class to generate the price associated with each unit of service in each user class, for example the volumetric prices per metric cube water demand in the residential class. Using the BCE residential users will pay higher volumetric rates, relative to industry, as the residential class will be assigned a larger proportion of the extra-capacity costs. A more detailed mathematical description of the BCE can be found in Appendix 6.B or in the AWWA manual.

The commodity demand method is the standard engineering rate setting method for utilities (see Kahn, 1988) and is similar to the base-extra capacity method but differs in the types of rate method classes used, allocating costs and service units by commodity and capacity, rather than by base and extra capacity. The commodity class refers to variable operating expenditures, while the capacity cost refers to all other capital and operating costs attributable to water demand.

These types of “cost of service” utility rates are generally criticized by the economic community (Bonbright, 1961; Hirshleifer, 1960; Kahn, 1988), which has suggested their replacement by some form of marginal cost pricing. The two principal critiques of fully allocated rate methods are the treatment of sunk cost within the volumetric charge, and the allocation of capacity costs among users on the basis of subjective criteria. The use of sunk costs in the rates is inefficient relative to marginal costs as these costs are previous costs that cannot be altered by current behaviour. Their incorporation in the volumetric charge leads to a dynamic pattern of pricing where volumetric prices rise after major capacity expansions (due to additional debt charges) and diminish over time as new capacity increments come into the planning horizon, the opposite to the dynamic, efficient pattern of prices suggested in section 6.3.

The second critique stems from the complete allocation of current capital or capacity costs of the utility among the different user classes on the basis of equity, thus leading to potential cross-subsidization between users. This emphasis on equity leads to a profusion of potential allocation formulas, including the industry standard methods described above. Both of these considerations serve to generate a price for the individual consumer that is different from the “escapable” costs (Lewis, 1946) that the consumer can affect by modifying his behaviour. Thus incorrect information is conveyed to the consumer on the consequences of his consumption, leading to a sub-optimal allocation of resources in society. In the case of the BEC, for example, peak demands are not

associated with residential consumers per se, but rather with all users using the system during the summer peak.

However, the base-extra capacity and commodity demand methods generate the rather interesting result that user classes with a more inelastic demand (e.g. residential consumers) are charged higher prices, and user classes with a more elastic demand (commercial and industrial classes) lower prices, through a declining block rate. Although this form of price discrimination does have a totally different methodology and rationale from Ramsey and Pareto Superior Nonlinear Output schedules, the practical structures are similar. Interestingly, (Kim, 1995) also found empirical evidence of U.S. utility rates approximating Ramsey price discrimination.

6.4.2 CWWA Manual

The CWWA manual integrates economic theory and current utility practice in developing a practical water rate. The rationale is principally based on efficiency, rather than equity, as well as improved practicality and rate stability. The rate method was developed as a conscious practical application of economic theory. The water rate is based on various estimates of marginal costs, with capital costs based on the present value of a stream of future investments, derived from the utilities capital plan. The theoretical basis for defining LRMC can be found in the preceding section on "Turvey Cost". The rate structure consists of non-linear pricing through a three-part tariff consisting of volumetric, connection and access charges. The volumetric charge is an estimate of long run marginal capital costs where:

$$(6.11) \text{LRMC} = \text{MCC} + \text{SRMC}.$$

MCC represents marginal capital cost and SRMC refers to short run marginal cost. MCC is approximated by:

$$(6.12) \text{MCC}_t = \sum^t \text{prVCC}_t / Q_t$$

Where VCC = Variable Capital Costs, Q is expected demand, t represents the rate setting and capital planning time period. r represents the amortization factor, and p the discount factor. Thus r in equation 6.18 above becomes:

$$(6.13) r = 1(1+i)^t / (1+i)^t - 1$$

Where i is the interest rate and p represents the standard discount factor, with d the discount rate;

$$(6.14) p = 1 / (1+d)^t$$

An option for peak load pricing is also presented, based on a seasonal summer peak where MCC is allocated to the summer rates.

Connection charges are used to recover the marginal capital costs associated with connecting to the distribution network. They are calculated according to current utility practice. They consist of two types; those related to connecting to the local distribution network, and those related to fixed annual operating costs (meter reading and maintenance, billing, etc.). The local distribution network charges are calculated outside of the annual water rates, while the annual operating charges form, along with the access charge, the short run fixed component of the annual water rate.

Access charge are used to recover the residual common costs that are not recovered through the volumetric or connection charges. Calculating the access charge requires an allocation of residual common costs to each connection based on equity, constrained by a limit-pricing rule. The limit-pricing rule suggested is that an access charge should neither cause a connection to connect to, or disconnect from the system. An example of an equity-based method to allocate the residual common costs is presented which is similar to current engineering rate setting practice as suggested for the annual connection charges. The equity-based consideration suggested to allocate residual common costs is previous water demand.

Accordingly, the CWWA method results in a two-part tariff, reminiscent of the earlier work by Coase outlined in Section 6.4.1. The first part of the tariff recovers the fixed portions (i.e., the connection and the access charges) of the utility's annual costs, and is constant for all users. The second part recovers the variable, or marginal, costs of the operation, including a forward-looking charge for future capital expenditure. This implies the existence of a long term capital plan for the utility, an instrument required in any event for effective operation and planning.

6.4.3 Summary

The rate setting methods outlined here diverge widely in their economic characteristics, data classification requirements, and length of use. The two AWWA methods are based on the average costs of service, and on the allocation of all costs to users on the basis of an essentially arbitrary rationale. The result in a type price discrimination among users that is economically sub-optimal, despite the fact that the DBR rate structure does correspond, in part, to the results obtained using Ramsay pricing methods. The AWWA methods require detailed cost accounting, and allocation of these costs among user classes, as noted somewhat arbitrarily.

The more recent CWWA method has marginal cost pricing as its theoretical basis, thereby being more justifiable in economic terms. The two-part tariff requires cost classification into fixed and variable components only, a procedure that we believe requires less effort, not more, as has been claimed. The key feature the CWWA method is the incorporation of long-term capital planning systematically into the rate-setting calculus. This is a key advantage over the long run because of the ability to influence capital spending through rate structuring. In economic terms, the CWWA methods has distinct advantages for future expenditures by water utilities

6.5 Water Pricing in Ontario: A Brief Overview⁶⁸

Water rates, and their resulting prices, are the single-most important contact that individual consumers have with water utilities, and the only direct signal they ever receive as to the cost of their use of public water resources. If these signals are incorrect, resource use will be distorted in a number of ways, such as excessive levels of demand, capitalization, and pollutant generation. Thus, it is important here to examine the issues of rate structuring and pricing

6.5.1 Rate Structuring

Four major types of water rates were used in Ontario over the study decade: flat rates; constant unit charges; and declining and increasing block rates (Table 6.2). Flat rates provide users with unlimited access to public water services in exchange for a periodic fixed amount of money. For example, if a given municipality's flat rate is twenty dollars per month, for that payment the

⁶⁸ The issue of water pricing is discussed in detail in a companion study commissioned by SuperBuild, and led by PriceWaterhouseCoopers (PwC) (Study 4). The authors of the present report are also working on the PwC study. This section is a brief summary of the latter, and is included here for contextual purposes.

customer would secure unlimited access to public water services. The marginal cost of water in this case is zero, which, of course, cannot be true in practice.

Pricing based on constant unit charges (CUC) operate in a manner similar to those used for purchasing most other goods and services. For each unit of water (e.g., a cubic meter) used, the customer pays a constant charge per cubic meter. Marginal cost of water in this case is equal to the unit charge, and the constant unit charging system is the one that could be most readily adapted to marginal cost pricing.

Block rate pricing systems are based on the concept that water usage is divided into a series of bands or blocks of varying widths. Water prices are set to vary according to the block (or blocks) into which any particular customer's water use falls. If the rates charged within the blocks fall over the range of water use, the rate type is termed a decreasing block rate (DBR) structure. The marginal cost of water under this system declines. Conversely, if the rates rise within the blocks over the range of water use, the rate type is termed an increasing block rate (IBR) structure, with increasing marginal costs. These changes in marginal cost will be "lumpy", in response to the blocking characteristic of both of these rate structures. Both types of block rate structures are economically inefficient because the fact marginal costs for the production and distribution of water are essentially constant. Block rates imply that these costs vary substantially and critically among users.

Table 6.2 Number and Percentage of Volumetric and Flat Rates, by Population Size Group										
Size Group ('000 persons)	Volumetric Rates, by Type								Total Residential Rates	
	Constant Unit Charges		Declining Block Rates		Increasing Block Rates		Flat Rates			
	Number	% of Total	Number	% of Total	Number	% of Total	Number	% of Total	Number	% of Total
1989										
1 - 5	16	37%	6	14%	0	0%	21	49%	43	100%
5 - 10	18	24%	15	20%	0	0%	42	56%	75	100%
10 - 50	21	23%	46	51%	0	0%	23	26%	90	100%
50 - 100	7	21%	3	9%	2	6%	21	64%	33	100%
>100	11	46%	8	33%	0	0%	5	21%	24	100%
Total	73	28%	78	29%	2	1%	112	42%	265	100%
1991										
1 - 5	28	23%	14	11%	3	2%	78	63%	123	100%
5 - 10	26	29%	17	19%	3	3%	43	48%	89	100%
10 - 50	39	33%	28	23%	7	6%	46	38%	120	100%
50 - 100	7	33%	8	38%	0	0%	6	29%	21	100%
>100	16	53%	5	17%	1	3%	8	27%	30	100%
Total	116	30%	72	19%	14	4%	181	47%	383	100%
1994										
1 - 5	38	28%	12	9%	5	4%	81	60%	136	100%
5 - 10	30	35%	14	16%	6	7%	35	41%	85	100%
10 - 50	38	36%	30	28%	8	8%	30	28%	106	100%
50 - 100	8	7%	6	5%	0	0%	105	88%	119	100%
>100	18	62%	2	7%	1	3%	8	28%	29	100%
Total	132	28%	64	13%	20	4%	259	55%	475	100%
1996										
1 - 5	36	27%	11	8%	8	6%	80	59%	135	100%
5 - 10	31	36%	19	22%	3	4%	32	38%	85	100%
10 - 50	49	44%	23	21%	12	11%	28	25%	112	100%
50 - 100	7	33%	9	43%	0	0%	5	24%	21	100%
>100	20	67%	2	7%	1	3%	7	23%	30	100%
Total	143	37%	64	17%	24	6%	152	40%	383	100%
1999										
1 - 5	18	22%	6	7%	2	2%	55	68%	81	100%

5 - 10	17	35%	5	10%	4	8%	23	47%	49	100%
10 - 50	54	39%	30	22%	10	7%	45	32%	139	100%
50 - 100	12	36%	7	21%	7	21%	7	21%	33	100%
>100	18	45%	10	25%	0	0%	12	30%	40	100%
Total	119	35%	58	17%	23	7%	142	42%	342	100%

Source: Data prepared for SBC study #4 by PwC.

Rate types are important because they act as implicit incentives for excessive or conserving water use practices⁶⁹. Here, it is conservation as a cost-saving concept, not a “principle of ethics that is important.”⁷⁰ For example, if customers have unlimited access to public water systems in exchange for a fixed payment per month (i.e., as under flat rates), they have no incentive to conserve on the use of water services. Under DBR structures, customers have limited incentives for conservation because they pay for water on the basis of volumes used. However under DBR structures, the incentive for conserving on water usage declines as larger and larger volumes of water are used. Under the CUC rate systems, the incentive for lowering water use is greater than under the declining block rate system because customers are paying the same amount for each unit of usage. Finally, under IBR structures, customers have an increasing incentive to conserve water services as they use increasing amounts of water, but, as already shown, these types of rate structures are economically inefficient. In summary, there is a direct linkage between the type of rate structure use and incentives for decreasing water demand.

Table 6.2 also indicates that, on an aggregate basis for the 1989-1999 period, between 40 and 55 percent of residential rates used in Ontario are flat rates. Flat rates are somewhat more common in smaller communities (population size groups 1 and 2, although there are exceptions, such as in the third municipal size range (i.e., 50,000 to 100,000 persons) in 1994. Declining block rates account for around between 13 and 29 percent of the rate structures used in Ontario during the decade of study. This figure was higher for 1989 because of the sampling procedures used. The remaining rate structures are accounted for by constant unit charges (principally) and increasing block rates (in a small number of cases). Interpreting these findings in terms of the demand for water, between 53 and 84 percent of residential water rate structures used in Ontario in the 1989-1999 period either had no incentive or decreasing incentives for conserving on the use of water. This may be a significant factor in increasing the amount of capital required for water utilities, both on the water supply side and waste treatment side of utility operations.

6.5.2 Water Prices

Water rates govern the level of water prices. Water prices to all users of municipal services are quite low in relation to the prices of most other goods and services. Table 6.3 illustrates this point in detail, using data from 1991, 1996 and 1999. These data show that average water prices are very low. To put these into context, Table 6.4 shows average prices as compared to the prices of other commonly used liquids. While this table appears slightly whimsical, it illustrates a serious issue in everyday terms. The table is compiled only for 1992, and therefore underestimate price levels faced in 1999. However, the general point is both clear and still relevant currently: water services in Ontario (as well as in most other areas [OECD, 1999]) are exceptionally cheap relative to other prices. This situation is common throughout the world, and illustrates succinctly a major part of the financing problems of water utilities.

⁶⁹ Water conservation may be perceived as an end in itself, as illustrated in section 2.1. However, it also has an economic interpretation, as a “socially beneficial reduction in water demand or consumption” (Baumann et al., 1979). It is in the latter context that the term is employed here.

⁷⁰ We do not dismiss the ethical connotation, but note that its assessment is beyond the scope of economics other than in a benefit:cost sense. This was addressed briefly in the demand management section of Chapter Four

Use Sector	1991		1996		1999	
	Total	Average/m ³	Total	Average/m ³	Total	Average/m ³
Residential – 25 m ³ per month	23.97	0.96	30.33	1.21	32.33	1.29
Commercial- 35 m ³ per month	35.62	1.02	42.03	1.20	44.80	1.78
Industrial – volume unspecified	n.a.	0.57	n.a.	0.70	Data not available	

Source: Data prepared for SBC study #4 by PwC.

To put these prices into context, Table 6.4 shows averages prices as compared to the prices of other commonly used liquids. While this table appears slightly whimsical, it illustrates a serious issue in everyday terms. The table is compiled only for 1992, and therefore underestimates price levels faced in 1999. However, the general point is both clear and still relevant currently: water services in Ontario (as well as in most other areas [OECD, 1999]) are exceptionally cheap relative to other prices. This situation is common throughout the world, and illustrates succinctly part of the financing problems faced water utilities.

Table 6.4 Typical Prices for Popular Liquids	
Beverage	Price (\$) ¹
Tap Water	0.96
Cola	805.00
Milk	985.00
Bottled/Mineral Water	1,500.00
Beer	2,500.00
Wine	9,000.00
Whiskey, Gin ...	26,700.00
¹ All prices are in 1992 Canadian dollars.	
² Only tap water includes the costs of automatic delivery to the user. The price includes all relevant waste treatment charges.	
Source: Tate and Lacelle, 1995	

6.6 An Overview of Empirical Studies Related to Future Water System Costs in Ontario

Although this report is largely theoretical in nature, it deals with the very real and expensive issue of providing adequate water servicing throughout the province. In this report, the most concrete manifestation of this issue rests in the costs of providing these services. The two principal aims of this section are to provide an overview of current costs of providing these services, and to review recent studies attempting to estimate future costs. While other SuperBuild studies are dealing with these issues in more detail, this review provides a useful picture of the size of the problems to which updated or new economic practices must apply. In the absence of the completed SuperBuild financial studies, we have abstracted some of the material presented here from the recent Delcan (2001) report commissioned by the Walkerton Inquiry.

The Government of Ontario has jurisdiction over the provision of water supply and waste treatment throughout the province, a jurisdiction that was delegated during the 1990s to the municipalities. The province can define the ways in which the costs for water services are recovered, and monitors these finances as part of the Financial Information Returns submitted annually by all municipalities. These FIRs provide comprehensive data on revenues and costs for all municipal functions, including segregated accounts on water servicing. These FIRs are the major source of data for this section of the discussion.⁷¹

6.6.1 Current Annual Expenditure

In 1997, the latest year for which completed data are available, expenditures on water and wastewater servicing totalled \$1.77 billion (Table 6.5). This amount was split almost evenly between water supply (52%) and waste treatment (48%). Inter-functional transfers refer to expenditures related to overhead within the municipalities, while other transfers pertain to charges by external organizations. Debt loads for water treatment was double that for water supply. Capital investment from current revenues (debt + transfers to own funds [i.e., reserve funds]) accounted for 48% of wastewater expenditures, but only 41% of water supply revenues. The bulk of this difference is accounted for by the higher labour-related costs for water supply.

⁷¹ This section is descriptive only, and does not include any description of financing or accounting principles. We assume that these will be adequately covered in other SuperBuild studies.

Table 6.5 Water Servicing Revenue Fund Expenditures (\$million) for Ontario, 1997			
Expenditure Category	Water Supply		Wastewater Collection/ Treatment
Salaries, wages and employee benefits	191	(20%)	133 (15%)
Net long term debt charges	78	(9%)	152 (18%)
Materials, services, rents, and financial expenses	285	(31%)	271(32%)
Transfers to own funds	290	(32%)	258 (30%)
Other transfers	8	(1%)	9 (1%)
Inter-functional transfers	58	(6%)	38 (4%)
TOTAL	910		861

Source: Ontario Ministry of Municipal Affairs and Housing, FIR database. Figures may not add due to rounding

Table 6.6 provides a slightly different view of water-related expenditures, by including capital fund expenditures (\$426 million for water supply; \$496 million for water treatment. Thus, cash outlays totalled \$969 million for water supply and \$947 for waste treatment. Of these amounts, 44% of water supply expenditures were for new investment; the corresponding percentage for waste treatment was 52%.

Table 6.6 Water Servicing Cash investment (\$million) for Ontario, 1997			
Expenditure Category	Water Supply		Wastewater Collection/ Treatment
Salaries, wages and employee benefits	191	(20%)	133 (15%)
Materials, services, rents, and financial expenses	285	(29%)	271(32%)
Other transfers	8	(1%)	9 (1%)
Inter-functional transfers	58	(6%)	38 (4%)
Capital fund	425	(44%)	496 (52%)
TOTAL	969		946

Source: Ontario Ministry of Municipal Affairs and Housing, FIR database. Figures may not add due to rounding

Table 6.7 indicates the sources of revenue for water servicing in 1997. In this respect, the Delcan (2001) report states:

“The outside contributions by means of grants are not large and have been decreasing. It is interesting to note that there is a greater reliance on property taxes for sewer system costs, at 12% of revenues. This is likely due to sewer costs historically being recovered from property taxes. A shift towards a sewer surcharge occurred in the 1970s when the regions were formed and they chose to move to a more "user pay" approach. It appears that the transition is still not complete. There is actually some justification for including some water costs on the property tax. Many municipalities charge the water system costs for providing fire protection to property taxes, also a legitimate approach supported by the AWWA and allowed in provincial legislation. There is no parallel for sewers.

The analysis of 1997 revenue sources indicates that fully 96% of water revenues and 95% of sewer revenues are from local sources. Only \$38 million or 4% of water revenues and \$45 million of sewer revenues came in the form of grants from outside sources. Thus most of the costs are locally funded.

Whether or not sufficient investment is currently being made in municipal water systems may be questioned.⁷² However, the recovery of current investment levels is very close to full cost recovery. The concept of recovering costs as much as possible through user rates is often promoted. Advantages include the promotion of conservation and clearly visible water and wastewater system costs. However, there are other revenue sources that are legitimate user pay methods of cost recovery. Capital costs are often recovered up-front for new servicing through frontage and connection charges, development charges, and contributions by developers. Also there are other fees and charge revenues reported that are levied based on services rendered. Thus it should not be assumed that the user rates should be carrying the total burden for water and wastewater costs.”

Source	Water Supply	Waste Collection/ Treatment
User fees	777	697
Other local	153	205
Outside Contributions	38	45
Total	969	947

Source: Ontario MMAH FIR database, as recorded by Delcan, 2001

The bolded sentence above is a key one. Water servicing costs certainly appear to be borne locally, suggesting the full cost pricing is already largely in place in Ontario. But, the crux of the issue is whether these costs are sufficient to meet current and expected needs. Data presented above, and various instances of water pollution, and the information given below suggest that these are considerable revenue shortfalls occurring in the system as a whole. Thus, the impression conveyed by Table 6.7, although based on the FIR data, may present a misleading impression about full cost recovery for water services in Ontario resulting from the accounting nature of the data presented.

6.6.2 Estimates of Future Needs

The estimation of future capital and O&M requirements for water and waste treatment are very difficult to make because of the large number of assumptions required, and the levels of servicing to be achieved. The latter problem is particularly serious with regard to waste treatment. Nonetheless, a number of attempts were made to do this during the 1990s, and it is useful to review these here. In doing this, it should be understood that the following estimates are very tentative, and to carry out an accurate, authoritative study requires a case-by-case approach.

a. Study #1 – Tate and Lacelle, 1995⁷³

This study formed part an examination of municipal water pricing in Canadian municipalities. The main interest was to obtain a measure of the impact of water service finance requirements on water prices to consumers. This study’s base year was 1993.

There were three basic phases in this study: (a) the calculation of water rate revenue based on the results of the 1991 Environment Canada water pricing survey; (b) and estimation of national capital + O&M requirements; and (c) analysis of additional revenues required via water rates. All of the calculations were made in terms of constant 1993 dollars; this obviated the need for financing assumptions. Phase (a) was calculated from the water rate data gathered from a national survey of water rates on municipalities with populations in excess of 1,000 persons, verified for Ontario using the FIR data base. Total estimated revenue from water rates in 1989 was \$1.18 billion, using the assumptions of the last paragraph. Independently, the FIR data indicated water rate revenues totalling \$1.1 billion. This confirmed the accuracy of the method used, and, accordingly, it was used

⁷² Bolding added.

⁷³ All cost data for this study were done in terms of 1992 dollars. No financing charges were incorporated

for Canada as a whole for 1991. Total estimated revenue was \$3.3 billion. Ontario accounted for 45% of this total.

MacLaren (1985) had estimated the total replacement value of Canadian municipal water and wastewater utilities at \$110 billion. There appeared at the time to be no more up-to-date country-wide estimate of water system capital costs, and the assumption was made that this estimate was "in the ballpark" for total required new capital expenditure. The average life of water system components, allowing for regular O&M, was taken to be 40 years, implying that \$2.75 billion should be spent annually on system upgrading. Considering a possible backlog in upgrading and renovation projects, it was assumed that 5% of the total replacement value would comprise the required expenditure for the 10 years from 1993 to 2003 (Table 6.8). To this must be added an allowance to install universal metering (estimated above at \$660 million). It was assumed that full metering would be done over the next 10 years. When all capital expenses are taken into account, annual total capital expenditures are estimated at \$5.66 billion for the period 1993-2003; \$2.75 billion thereafter. From this estimate, the "regularly scheduled" expenditures of \$1.8 billion were deducted. Thus, over the long run, net new capital outlays total \$3.76 billion for the 1993-2003 period; \$0.95 billion thereafter (Table 6.8).

Table 6.8 Summary of Estimated Water System Revenues and Costs (\$ billion, 1993) to Achieve Water Servicing System Adequacy				
Costing Element	Total Cost	Annualize Revenue/ Cost, 1993-2003 ¹	Annualize Revenue/ Cost, 2003 →	Source
Annual revenue	n.a.	3.3	3.3	Table 2.14
Total replacement value	110	n.a.	n.a.	MacLaren (1985)
Annual capital cost for system components	n.a.	5.5	2.75	40 year average replacement period for \$110 billion of capital assets
Capital outlay for meters	0.56	0.06	n.a.	3.3 million meters @ \$200 each, done over 10 years.
Total annual capital	n.a.	5.56	2.75	
Current capital outlays	n.a.	1.80	1.80	Estimated from telephone survey
Net capital requirement	n.a.	3.76	0.95	Calculated
New annual O&M	n.a.	0.83	0.83	15% of new capital for the 1993-2003 period (assumed).
Total annual net new money requirements	n.a.	4.59	1.78	Calculated

Source: Tate and Lacelle, 1995.
¹1993-2003 was considered as a "Catch-up" period to allow for an existing backlog of work.

Additional Revenue Sources Using Water Pricing

The research continued to examine ways in which raising these financial requirements, in the context of water rates, as the primary vehicle for acquiring needed revenue. If the rate structures remained as they were in 1991 and price levels doubled on average nationally (roughly an additional \$20/month per connection), an additional \$2.2 billion could be raised. This would clearly meet the monetary requirements beyond 2003, but would be inadequate for the 1993-2003 period. A number of options (Table 6.9) were examined for raising the revenue stream to the required level. Two possible options were modeled in the study:

- An 80% sewer charge⁷⁴ in municipalities that currently have no sewer charges;
- The adoption of full metering of residential customers. Two pricing options were tested following full metering: adoption of the regional average monthly water price for newly metered customers between the 25 m³ and 35 m³ levels of usage; and, where the latter were less than the average national monthly price, adoption of the latter price.

Region	Doubling 1991 Rate Levels (adjusted for price elasticity)	Adding an 80% Sewer Charge in Municipalities Without Them	Adopting Full Metering for Residential and Commercial Services and Charging Regional Average Price	Total Additional Revenue
Atlantic	131.6	98.3	10.2	240.1
Quebec	305.7	388.7	161.7	856.1
Ontario	1,015.5	852.1	44.9	1,912.5
Prairies	597.2	502.7	24.0	1,123.9
B.C. + Terr.	182.3	149.4	77.3	409.0
Canada	2232.3	1991.2	318.8	4,541.6

In analyzing the revenue effects of water price increases, account has to be taken of the price elasticity of demand. Decreased water demand can be expected to result from increases in the price of water (see Chapter Four). Water demand is quite price inelastic, in that a given percentage increase in price will lead to a less-than-proportional decrease in water usage. Tate and Lacelle assumed an average price elasticity of -0.2, which implies, for example, that a 10% increase in price will lead to a 2% fall in water demand.⁷⁵ Thus, for a doubling of basic water prices will result in a 20% decrease in usage. Similarly, the 80% sewer surcharge will cause a 16% decrease in demand. Accordingly, the pricing modifications proposed in the simulation analyzed here will lead to an estimated demand decrease of 32%. This effect has been incorporated into Table 6.9.

In addition to the \$2.2 billion that could be raised through a general doubling of water prices, a country-wide total of \$2.0 billion would be raised through a 80% sewer charge. (Sewer charges of this magnitude are already in use in some municipalities.) Universal metering and pricing reform in flat-rate communities would raise between \$0.1 billion and \$0.3. As shown in Table 6.9, additional annual revenues for the complete simulation would total between \$4.3 billion and \$4.5 billion, depending upon whether the metering option was adopted. These fairly inexpensive water rate and pricing reforms would meet the financial requirements for both the "catch-up" period (1993-2003) and the period beyond 2003. Thus, at the macro-economic level all of the forecast financial requirements in Table 6.8 could be met through the pricing reforms simulated here. possible options were modeled in the study: possible options were modeled in the study:

b. Study #2 - Ontario Ministry of the Environment, 1996

A discussion paper by OMEE estimated the future costs for water servicing works. The paper appears to be based on a detailed analysis of the state of water and wastewater facilities in the

⁷⁴ This percentage reflected the "typical" Canadian practice for municipalities using sewer charges at the time the study was conducted.

⁷⁵ As shown in Chapter Four, this elasticity value may have been on the low side of the empirical range.

province. The findings are shown in Table 6.10. The paper did not deal with water distribution of sewerage systems.

The study placed these monetary requirements into the context of full cost recovery, using FIR data for 1993. The difference between rate-based revenues and revenue fund expenditures totalled \$435 million (Table 6.11), which was met through various means, such as property tax revenue, reserve funds, transfers, etc. The interesting feature of this table is that it identifies an approximate 25% shortfall in rate revenues in terms of expenditures. The OMEE paper examined means of meeting the shortfall of current revenue (i.e. \$435 million), plus the required new expenditure of \$6,045 million. Unlike the Tate-Lacelle paper, OMEE used a 40-year financing period with a 7.75% interest rate. Results are given in Table 6.12. This study shows a 73% increase of water rates to cover these shortfalls. It assumes, as we also believe, that water revenues should shift to the user and away from tax or frontage based charges. It does not take account of possible water conservation impacts of increased water rates, as was done in the Tate-Lacelle study.

Table 6.10 Capital Requirements (\$ 10⁶) for Water Supply and Waste Treatment, 1995-2005				
System Component	Deficiencies	Rehabilitation	Growth	Total
Water Supply Plants				
- Overcoming capacity deficiencies	329	n.a.	n.a.	329
- 1995 – 2000 planned investment	n.a.	376	229	605
- 2000 – 2005 planned investment	n.a.	535	376	911
Total Water Supply	329	911	605	1,845
Wastewater Treatment Plants				
- Upgrades of 21 primary-level plants to secondary	746	n.a.	n.a.	746
- Overcoming capacity deficiencies	595	n.a.	n.a.	595
- 1995 – 2000 planned investment	n.a.	785	492	1,277
- 2000 – 2005 planned investment	n.a.	911	671	1,582
Total Wastewater	1,341	1,696	1,163	4,200
Total Water System	1,670	2,607	1,768	6,045

Source: OMEE (1996) as interpreted by Delcan (2001).

Table 6.11 Water and Wastewater Revenue Fund Expenditures and Water Rate Revenue		
	Expenditure/Revenue – Total \$ million	Expenditure/Revenue - \$/m ³ , ¹
A. Revenue fund expenditures	1,707	0.93
B. User rate revenues	1,272	0.69
Difference (A – B)	435	0.24

Source: Delcan, 2001.

¹ The calculation of revenue per cubic meter was based on a total flow of 1.84 billion m³

Table 6.12 Using Full Cost Rates to Meet Water System Revenue Shortfalls		
	Expenditure/Revenue – Total \$ million	Expenditure/Revenue - \$/m ³
Water rate revenues	1,272	0.69
Current Revenue Shortfall	434	0.24
Amortized Water Supply Needs	342	0.19
Amortized Waste Treatment Needs	151	0.08
Total	2,200	1.20
% User rate increase		73%

c. **Study #3 – Canadian Water and Wastewater Association (CWWA). 1998**

This study was prepared by the CWWA, supported by Canada Mortgage and Housing Corporation (CMHC), to estimate water system investment needs in Ontario between 1997 and 2012. Standards

to be met were the Canadian Drinking Water Standards, separating storm and sanitary sewers, and Level III (tertiary) wastewater treatment

The study estimated annual water and sewer system expenditures of \$1.84 billion, and wastewater treatment annual costs of \$4.09 billion. The costs were in constant dollars, and thus total water and sewer expenditures were \$27.6 billion, waste treatment \$61.4 billion.

Environment Canada municipal survey data was used to show that 85% of the 1994 Ontario population (total = 10.9 million) living in urban places of greater than 1,000 persons. Of these, 92% were served by water supply, 90% by some form of waste treatment. The estimated length of new water mains required to serve the remaining population was 3,862 km, based on providing 193 km per capita. To meet the need of an increased 2.8 million persons (i.e., a 30% population growth) translated to an additional 14,424 km. of mains. Other assumptions made included:

- 0.6% of the existing system to be replaced annually;
- main extension cost of \$200 per meter; replacement and restoration cost of \$300 per meter;
- water supply expansion cost of \$2,000/capita; base upgrades of \$300/capita and major upgrades of \$400/capita.

The study results are shown in Table 6.13.

System Component	Replacement	Expansion	Growth	Total
Mains	1,164	1,495	2,885	5544
Storage	316	37	137	490
Supply & Treatment	1,100	1,803	1,142	6,529
Total	2,590	1,830	8,142	12,563
Annual Average	173	122	543	838

Source: CWWA 1998, as interpreted by Delcan, 2001

d. Study #4 Ontario Sewer and Watermain Contractors Association (OSWCA), 2000

This study was carried out by PriceWaterhouse Coopers for the OSWCA, primarily to investigate the concept of full cost pricing for Ontario, and to identify means for addressing any mitigation measures that might be required, should policy move in this direction. Data used were drawn from the CWWA study outlined above, augmented by the FIR database. The consultants appear to have accepted many of the CWWA procedures and assumptions, but lowered the sewer costs with respect to the separation of storm and sanitary sewers. For 1997, for example, the study found that water servicing rates would have to rise by 31% to meet the full cost recovery objective. The study included storm sewers in the user pay analysis, which is somewhat unusual, because these costs are not typically borne by water users, because there seems no equitable way of allocating these costs.

Results of this study are shown in Table 6.14. The overall impact of employing full cost pricing, according to this study, would be a 31% increase in average water rates.

Table 6.14 The Estimated Cost of Additional Water System Expenditure, 1997			
Cost Component	Water	Sewer ¹	Total
Revenues			
- Water rate	780	723	1503
- Fees and services charges	32	39	70
- Special charges	12	12	25
Total	824	774	1,598
Operating Costs			
- Salaries and wages	192	148	339
- Net long-term debt charges	78	171	249
- Materials, services, rents	288	289	577
- Other transfer	8	9	17
- Interfunctional transfers	58	44	102
Total	624	660	1284
Available for capital investment	200	114	314
Sustaining Capital Expenditures			
- Replacement	98	401	499
- Expansion	164	145	309
Total	262	545	807
Additional Funding Required	62	431	493
Required Water Rate Increase	8%	56%	31%
Source: Delcan, 2001			
¹ The sanitary sewer revenues total \$748 million, and operation costs are \$637 million, leaving a total "available for capital investment" of \$111 million. This is close to the \$114 million "Available for capital investment" reported in the table. Thus, the inclusion of storm water is apparently insignificant.			

e. Study #5 The Powell Study (2000)

This study by George Powell, a respected engineer in Ontario, was produced as a conference paper. The approach taken was somewhat related to that done by Tate and Lacelle, in that Powell started with a "macro-level" estimate of Ontario's water infrastructure of \$50 billion (\$35 billion in-ground and \$15 billion above-ground). He then estimated an "average life" for these assets. The average life for in-ground assets was taken as 75 years, that for above ground assets 35 years. Accordingly, an annual replacement cost of \$895 million would be required [(\$35 billion/75) + (\$15 billion/35)]. Powell also used an OMEE study from the early 1990s, which estimated the 15-year capital costs of water systems to be \$19 billion, or \$1.3 billion per year on average. The Powell study's findings are summarized in Table 6.15.

Table 6.15 Summary of Water Servicing Costs (\$ million) from the Powell Study				
Cost Component	Water Supply	Waste Treatment	Total Annual in constant dollars	Notes
Capital replacement	447.5	447.5	895	As in text
Meeting Safe Drinking Water Standards	133	n.a.	133	OMEE internal study – 15-year cost = \$2*10 ⁹
Metering	33	n.a.	33	OMEE internal study – 15-year cost = \$500*10 ⁶
System Rehabilitation	117	117	234	OMEE internal study – 15-year cost = \$3.5*10 ⁹
Waste Treatment	n.a.	900	900	OMEE internal study – 15-year cost = \$13.5*10 ⁹
Growth Allowance	50	50	100	Assumed by Powell
Total	780.5	1,514.5	2,295	
Source: Powell (2000), as interpreted by Delcan (2001)				

Summary of Cost Projection Studies

Table 6.16 illustrates the range of estimates to summarize the results of the 5 studied reviewed here. There appears to be little consistency in these estimates. All but one, however, places the costs of waste treatment higher than that for water supply.

Table 6.16 Comparison of Additional Annual Water System Expenditures \$10⁶) for Ontario Water Systems				
Study	Water Supply	Waste Treatment	Total	Comments
Tate and Lacelle (1995)	846	1034	1880	50% of national requirements (\$3.6 billion/year) allocated to Ontario. Supply:Treatment ratio = 45:55. See Table 6.8.
OMEE (1996)	184.5	420	604.5	Costs annualized using a 40year life for facilities. Distribution and collection costs NOT included. See Table 6.10.
CWWA (1998)	621	218	838	Total costs allocated equally over 15 years. "Supply" category reported by Delcan (2001) assumed to be 50% treatment costs. Costs in constant dollars. See Table 6.13.
OSWCA (2000)	62	431	493	See Table 6.14.
Powell (2000)	780	1,514	2,295	See Table 6.15.

6.7 Summary

This Chapter addressed the economic theory and common practice relative to water pricing in Ontario. Initially, the chapter develops a definition of full cost pricing which, being an economics paper, suggests three efficiency-based criteria for defining full cost recovery based on marginal cost pricing;

- peak load pricing;
- recovery of marginal distribution network costs through connection charges; and;
- forward looking volumetric rates, incorporating long run marginal cost pricing, and estimates of future, rather than sunk, capital costs;
- extra-strength sewer surcharges.

In terms of full recovery, the report suggests that certain types of subsidization would not compromise full cost recovery, with possibly efficient subsidies limited to either common costs deriving from sub-additivity in the cost function, or costs associated with externality mitigation, such as water quality or effluent quality upgrades.

The reconciling of economic full cost recovery solutions to the problem of economies of scale typically involves water utility cost functions with average costs exceeding marginal costs. This problem has resulted in several different theoretical suggestions as to how to achieve both prices based on marginal costs and full cost recovery. The main solutions to this revenue gap are proposed here in the chronological order in which they were published:

- subsidization by senior levels of government, and pricing at marginal costs

- Coase two part tariff: Volumetric pricing at marginal cost and imposing an additional fixed access fee on public utility users
- Ramsey pricing: Mark-ups on marginal costs based on charging users with inelastic demands higher prices than users with inelastic demands
- Pareto-optimal linear outlay schedules, or volume discounts for large users

Some discussion was also devoted to increasing block rates, which are methods for ensuring full-cost recovery in utilities generating excess profits due to diseconomies of scale, for example drought in water scarce areas. Increasing block rates can also function as a means of transferring revenues from large water users to smaller water users on equity grounds, and is often recommended on water conservation grounds.

In addition to these proposed rate systems, economic theory has developed one other major theoretical advanced based on efficiency maximization;

- peak load pricing, or seasonal summer marginal cost pricing, in the case of water utilities

This implementation of peak load pricing is assessed under the assumptions of sub-additivity in the long run production function, indivisible capital, exogenously increasing demand, and optimal capacity sized to meet peak demands. The analysis indicates that under these assumptions, and using the standard definition of costs, there is no continuous long run marginal cost function (LRMC), simply a succession of discontinuous short run marginal cost (SRMC) functions corresponding to optimal discrete capacity increments. Thus the peak and off peak prices correspond to the intersection of the respective demand curves and short run marginal costs, and off-peak prices may exceed peak prices, depending on the particular time period within the capacity cycle. In general, peak and off peak prices will be less than average cost, necessitating an additional access charge to achieve full-cost recovery.

Long run marginal cost is rehabilitated as a basis of rate-setting, with a discussion of some different methods for estimating LRMC, given discontinuous capital. The concept of Turvey cost, and the extension to it, generate a continuous capital cost function which approximates the movement in SRMC over time, rising prior to capacity increments and falling afterwards, but smoothed to reflect all discounted future capital costs over the long run planning horizon. From this perspective LRMC is defined as a discounted moving average of future capital expenditures.

All of the rate structures derived from economic theory are based on maximizing social welfare, or as discussed in the principles section in chapter two, Pareto optimality. However, only two can be considered to be “first best” solutions, or solutions which allow for prices equal to marginal costs, with these being the Coase two part tariff, under certain conditions, and peak load pricing. The condition for Pareto optimal pricing occurs only if the access charge does not induce any users to disconnect or not connect at all, to the utility grid. A first best solution requires that the marginal cost equal price, while second best solutions minimize the social costs associated with divergence from the marginal cost-pricing rule, in a static setting.

One of the rare rate setting manuals for water utilities incorporates marginal cost pricing is the manual for the Canadian Water and Wastewater Association (CWWA, 1993). This manual uses both the Coase two-part tariff and peak load pricing as the basis for their rate setting methodology.

In contrast, the prevalent rate setting methods in use in North America, and Ontario, are engineering methods based on average cost pricing, with cost allocation to user groups based primarily on equity criteria. The expressed rationale is that, because residential users have more pronounced peaking behaviour, it is fair that they should be allocated a larger share of the sunk or historical

capital costs, thereby generating the declining block rates commonly found in current rate setting practice. The two principal rate setting methods advocated by the AWWA are;

- the base extra capacity method, which allocates costs between base and “extra” capacity.
- the commodity demand method which allocates costs between commodity and capital costs.

The engineering rate setting methods described above have the interesting property of approximating Ramsey pricing, in generating cost recovery by charging users with higher demand elasticities lower prices.

Finally the chapter assess’ the degree of full cost recovery in Ontario and concludes that local municipalities are bearing most of the costs for water servicing. However, there is considerable doubt whether this local, user-based funding system is currently ensuring adequate economic and financial resources for the future. Thus, even though full cost pricing seems to be in place, this may be illusory, because prices and the related level of investment appear to be below the level required to upgrade and maintain current water infrastructure. Some empirical estimates of this revenue inadequacy are provided with annual estimates ranging from \$ 498 million to \$ 2.2 billion, depending on the type of upgrading specified.

Appendix 6.1: Capital Cost Accounting, Utility and Cash Based Accounts

Given the importance of capital in the water industry the means by which utilities account for capital costs will have an effect on the water rates, principally by determining the annual revenue requirement, or costs, that must be met from user charges. In the absence of subsidies from senior levels of government the annual revenue requirements will correspond to the annual total accounting costs incurred in providing services. In Canada and in the United States water utilities organize their capital accounting records using either a cash or utility-based system of accounts (also called respectively fund-based and fixed-asset accounts).

Utility-based accounts adhere to standard commercial accounting conventions in their treatment of capital costs. Investments are recorded as additions to fixed assets while annual depreciation allowances represent the loss of asset value over time, generally calculated by means of accepted commercial accounting definitions of depreciation (i.e. historical cost, straight line). For perpetual public utilities depreciation allowances are also used in conformance with respective tax laws in order to minimize taxes paid on commercial profits.

Utility-based accounts also track the sources of capital funds, namely debt and equity. Interest cost on debt enters the accounts as an annual cost, as does the return on equity determined by the appropriate regulatory authority. However, the repayment of principal is not recorded as a cost against revenues, since to do so would duplicate the capital cost that is already recorded as annual depreciation and would be double counting. Perpetual public utilities must also pay taxes on commercial profits earned.

Utility-based accounting is normally required for perpetual public utilities that are allowed to earn a return on investments. In Ontario, a publicly owned water utility may use a utility-based system of accounts in order to achieve a better understanding of the state of its finances, and to improve decisions in the management of its resources. However, utilities that do must also maintain a second set of accounts, cash accounts being required for municipal financial reporting purposes in Ontario.

Cash-based accounting corresponds to the method used for government accounting where a return on investment is not calculated, and taxes are not paid. A cash-based account uses separate funds to track revenues earmarked for specific purposes. The three basic types of funds are revenue funds, capital funds and reserve fund.

The revenue fund account is the operating account of the utility. Current expenditures are recorded in the revenue fund, as well as income from user fees, transfers from general revenues, and possibly other sources such as investment income and grants or transfers from senior governments. Revenue fund expenditures include operating costs, debt service charges, savings that are transferred to reserve funds, and any capital costs that are paid from current revenues.

Capital fund accounts record capital expenditures or investments. Money flows into a capital fund from credit sources (debentures, credit notes), as transfers from revenue and reserve funds, or as grants. Revenues are accumulated over time to cover future capital investments and assigned to reserve funds until they are transferred to capital funds to be used.

Capital costs are represented in a cash-based accounting system by the accumulation of revenues in reserve funds, the assumption of debt, the expenditure of funds out of capital fund accounts, and by any related debt service costs (payments of interest and principal) that appear in the revenue fund account.

A cash accounting system does not report fixed assets and accumulated depreciation and it does not allow annual depreciation and a return on equity to be identified as costs in the revenue fund. Costs which appear in cash accounts in connection with a particular investment will depend on how a municipality chooses to use reserves, current revenues and debt to finance that investment.

Cash-based accounts measure the actual expenditures and cash flow taking place within a utility, while utility based accounts represent the additions and debits to a stock of capital assets, measured in commercial accounting terms. In general, in any one year the two systems will generate different estimates of the total annual revenue requirements, with the difference being more pronounced the greater the difference between the financing measures used (e.g. debt, transfers, equity) and the greater the disparity between amortization of government capital investments, and depreciation in utility based accounting.

While utility-based accounting is appropriate for reporting to a regulatory body (or to stockholders), which must impose some form of consistent estimate of utility revenue earning potential, capital fund accounting is more useful to a government utility which does not earn a return on equity, nor pay taxes, but which must recover its cash needs through its water rates. The issue of cash or cost based accounting is relevant to full-cost recovery as annual revenue requirements will differ depending on which accounting system is used.

Appendix 6.2: The Base-Extra Capacity Method

The base extra-capacity method involves allocating current system costs by three different cost classifications, by object, by function and by rate method class. The functional classification allocates costs by the system component (e.g. treatment plant, distribution mains), while the object classification allocates costs by accounting method (e.g. depreciation, return on rate base, operating and maintenance expenditures).

The rate method class is a cost classification peculiar to the rate method⁷⁶, in the case of the base-extra capacity method; base costs, extra capacity costs, customer service costs and fire protection costs. The extra capacity costs are further subdivided into maximum daily demand and maximum hourly demand costs, while customer service costs are divided between meters and service costs, and billing costs. The volumetric charge is based on the base and extra capacity costs, the fixed charge on the customer service costs, and the fire protection costs are recovered from property taxes.

To place the discussion now in more abstract and mathematical terms, let $i=(1,,f)$ index the number of costs allocated by both functional and object categories, and let $j=(1,,m)$ index the number of rate method classes. Let C_i denote a cost allocated by object and function, and C_{ij} a cost allocated by object, function and rate method class.

Following the allocation of costs service units are allocated to each user class (e.g. residential, commercial, industrial) in order to calculate the unit cost of service per rate method class. The units of service used in calculating the unit cost of service for base and extra capacity costs are volumetric measures (e.g. thousands of gallons per day), those for metering are based on meter capacity, while those for billing are based on bills issued. As mentioned above fire protection costs are billed separately through property taxes; thus no units of service are calculated.

Let $l=(1,,t)$ index the number of types of service units, and let $k=(1,,u)$ index the number of user classes. Let S_l denote units of service by service type (e.g. thousands of gallons per day), S_{lk} denote units of service by service type, by user class (e.g. thousands of gallons per day, per residential class) and S_{ljk} denote units of service by service type, by user class, and by rate method class (e.g. base and extra capacity). The last allocation is accomplished through the use of a proportional parameter (a) such that;

$$(6.13) S_{ljk} = a_{kj} * S_{lk}$$

Where,

$$(6.14) 1 = \sum_{k=1}^m a_{kj}$$

The determination of respective parameters for allocating the units of service is theoretically based on a detailed examination of previous peaking behaviour in the different user classes. However, given the lack of sufficient metering for accurate estimates the manual accepts a large measure of discretion and value judgement on the part of the rate-maker in allocating service units to different user classes.

⁷⁶ The other rate method (commodity demand) uses commodity (variable operating) and capital costs as rate method classes.

The unit costs of service per service type and rate class are calculated by;

$$(6.15) \text{ UCS}_{ij} = \frac{SC_{ij}^f}{SS_{ikj}^u}$$

Where UCS = unit cost of service. In order to calculate the total cost per user class, per rate method class;

$$(6.16) \text{ TC}_{kj} = \sum_{ij}^t \text{SUCS}_{ij}^* \text{S}_{ikj}$$

Where TC equals total cost.

The final step in determining the volumetric charge per user class is to sum the total costs per user class in the base and extra method cost classes, then divide by the expected annual water use of the appropriate user class. The different prices for each user class are then structured into a declining block rate where residential consumers fall into the first and highest block, commercial users into a lower block, and industrial consumers fall into the lowest price blocks. The residential consumers fall into the initial and higher block rates due to their more pronounced peaking behaviour, and thus higher proportion of the extra capacity costs, while industrial and commercial clients fall into lower blocks due to their more even peaking behaviour.

The fixed charge is calculated by appropriately allocating the total costs for metering and billing to each connection served by the utility.

Appendix 6.3: Mathematical Appendix to Chapter 6

A6.3.1 Peak Load Pricing with Two Time Periods

(taken from Williamson, 1966);

Assume constant LRMC, and a constant SRMC up to some capacity limit, where SRMC becomes perfectly inelastic. Assume the plant is used to capacity only under the peak load.

Let social welfare be represented by;

$$(A6.1) \quad W = (TR_o S_o)w_o + (TR_p + S_p)w_p - bQ_o w_o - bQ_p w_p - kQ_p$$

Where p, and o refer to peak and off-peak time periods respectively, w refers to the respective weights or fractions of the two time periods over the whole, TR= to total revenues, S to consumer surplus, Q to quantity demanded, b to unit operating costs, and k to unit capital costs.

Differentiating W for each Q yields;

$$(A6.2) \quad P_o = b$$

$$(A6.3) \quad P_p = b + k/w_p$$

A6.2 Discontinuous Capital and Dynamic Programming

(taken from Manne, 1961, Scarato 1969)

Assume a constantly increasing inelastic deterministic demand, a Cobb-Douglas function for capital exhibiting economies, no depreciation, and long run constant returns to scale. Form the recursive equation;

$$(A6.4) \quad C(x) = k(x)^a e^{-rx} c(x)$$

Where x=increment of capital, a is the scale parameter, k is a cost constant, r is the interest rate, and C is cost.

This represents the Bellman equation, where the long run constant returns to scale assumption, and constant increasing demand indicates that each capacity increment will be the same size, and take place at the same time relative to the previous capacity increment. Thus assuming an infinite time horizon;

$$(A6.5) \quad \frac{C(x)}{k} = \frac{x^a}{1-e^{-rx}}$$

Taking logs of each side of equation 6.5 and differentiating for x generates;

$$(A6.6) \quad a = \frac{-rx}{(e^{-rx})} - 1$$

Where the optimal increment of capital is determined by the relationship between a and r.

Manne also solves for a similar function under the assumption of a random walk in demand with a similar result, but larger capital increments.

CHAPTER 7: WATER UTILITY ORGANIZATION AND MANAGEMENT

7.1 Introduction

This chapter examines water servicing systems as network utilities, and the ways in which the network utilities concept may provide alternative forms of organization and management. Network industries are generally characterized by some type of physical connection between their consumers and producers. In the case of water utilities, networks include the mains and sewers, storage facilities and pumping stations which distribute potable water to individual service connections, and sewage to the treatment plants or to receiving water bodies. Other examples of these types of industries include electrical utilities, natural gas utilities, telecommunication industries, and transport industries, such as railways and roads. These industries constitute what is often referred to as the physical infrastructure of an economy and society.

To reflect back on Chapter Three, for most of the 20th century, a substantial body of economic theory was developed on the unique nature of these types of industries, based on the concept of natural monopoly associated with economies of scale. Essentially, this theory was based on a single product firm with a declining average cost of production, resulting in the good (e.g. potable water) being most efficiently provided by a single firm. This conceptual model has been used throughout this report so far.

Two main forms of industrial organization have been traditionally associated with these industries: public ownership and regulated utilities, with the latter reporting to public bodies that regulated rates and service levels, generally in the form of rate of return regulations. The rationale for these forms of organization were the possible abuse of monopoly power in the form of higher rates, and the need to prevent lower service levels than would be socially optimal if these good were provided by unregulated private industry⁷⁷. Public ownership was, and continues to be, the principal form of ownership associated with water utilities in industrial economies, and certainly predominates in Ontario and Canada. The United States has exhibited a slightly higher tendency towards the use of regulated public utilities, primarily in smaller urban centres; however public ownership also predominates in that country.

In the late 1960s and early 1970s, however, many advances were made with respect to the sophistication of economic theory regarding network industries. Three of the most relevant were:

- the addition of the concepts of economies of scope and sub-additivity in assessing natural monopolies, and thus a greater emphasis on separating utility services, and introducing competition;
- the concept of contract bidding for the rights to build and/or operate utilities, and the associated considerations of transaction costs; and
- the recognition that the form of utility regulation has an impact on utility behaviour, and the development of incentive regulation

This chapter will review some economic theory related to water utility organization in three sections: a) introducing competition, b) transaction costs and c) incentive regulation. The empirical aspects of

⁷⁷ Both of these forms of market failure are characteristic of monopoly situations, either private or public, as demonstrated in Chapter Three

public policy will be addressed through studies on the impacts of alternative forms of utility management in England, France and the United States. The last section will address theory and practice regarding innovative financing mechanisms, using examples from the same countries.

7.2 Economic Theory

7.2.1 Introducing Competition

The rationale for utility restructuring has generally been that of improving efficiency through the introduction of competitive markets. The efficiency of markets for allocating goods has many justifications in economics, two common rationales are the static efficiency of marginal cost pricing, and dynamic efficiency of prices as information. The benefits from market competition are constrained somewhat in the case of water utilities by possible market power abuses (e.g., monopolistic effects), and the possible transaction costs involved in developing competitive conditions, as will be discussed in the next section. These problems are heightened by the unique role of water utilities in supporting public and environmental health.

Three main forms of competition that are important here pertain to output, input, and capital markets. Output competition refers to competition in the provision of goods (e.g., potable water), two water utilities competing to provide water to a single connection. The sub-additivity of the distribution network makes this an unlikely possibility for water utilities in the full retail market, for example, it is unlikely that cost efficiencies would result from building two or more competing distribution networks. However, some aspects of output competition do exist in wholesale markets for water, whereby separate treatment plants are both possible and common. For example, industrial water use and treatment at large plants in Ontario is self-supplied, rather than utility provided, while substantial portions of the residential population continue to use private wells. The wholesaling of water by one municipality to another is also common.

Input competition refers to competition in the production of the good. An example would be sub-contracting or out-sourcing different aspects of the construction and/or operation of water utilities through some form of competitive tendering process. The concept of franchise bidding for public utilities was advanced in economic theory by Demsetz (1968) in a classic paper called, "Why Regulate Utilities?". The paper suggested replacing rate regulation by a regular competitive bidding process between private firms for the contract rights to operate public utilities, thus generating improved incentives for cost efficiency and innovation. Franchise management of water utilities is the most common form of organization for water utilities in France, the western country with the most experience with private water companies, as well as in many Third World countries. It is also the most common form of utility organization in the small, but growing, trend towards public/private partnerships in Canadian water utilities. Indeed, of all utilities, "competitive tendering has been used most extensively as a regulatory tool in the water utilities, where the scope for output competition is limited, and therefore the pressure on costs to prevent the loss of market share is very weak." (Helm and Jenkinson, 1998).

Capital market competition refers to the possible discipline imposed by shareholders on publicly owned utilities to minimize costs. In competitive capital markets competing sets of owners and managers can take over the assets of under performing utilities, generating a stimulus for efficiency improvements. The problem with privatization, and capital markets, is that these same stakeholders may also induce socially unacceptable levels of service and profits, due to the public good and natural monopoly characteristics of water utilities. Thus, fully privatized, shareholder owned water utilities are almost invariably regulated by some form of government agency, typically for both water quality and quantity considerations, as well as in terms of economic regulation regarding acceptable levels of profits, costs and services. There are currently no privately owned

water utilities in Ontario, though a large provincial crown corporation (OCWA) has a dominant market share of operating and maintenance contracts.

While some writers also classify certain regulatory strategies (Helm and Jenkinson, 1998) as competition, it is more accurate to label these as competition mimicking strategies, or incentive regulation. Examples of incentive regulation in the water industry are price caps and benchmarking, as are used in the economic regulation of private water utilities in England.

While economic theory stresses efficiency gains from competition and free markets, a common limitation to these efficiency gains derives from inherent market power, either in the form of monopoly or oligopoly. A common complaint in the modern economics literature relative to introducing competition in formerly regulated monopoly markets is the tendency for the introduced competition to be restricted to fairly tight oligopoly markets. To quote from (Shepperd, 1998):

- both single firm market dominance and several firm market dominance (tight oligopoly) are usually inconsistent with effective competition;
- regulation can have important benefits. The supposedly severe costs of regulation have not proven to be large; and
- premature deregulation is harmful to society. It permits monopoly power to remain, but without the restraints necessary to protect the public.

It can be noted that the French, particularly, and U.S. markets for private water utilities contracts are fairly tight oligopolies, with the French market comprising essentially two firms (Suez, Vivendi). However, international competition for contracts are increasing as French and English companies penetrate foreign markets, both in the developed world, and more particularly in underdeveloped world markets.

A somewhat different perspective on efficiency stems from the importance of innovation. As Solow (1957) demonstrated, it is innovation that primarily accounts for economic growth in the long run, while economists such as Shumpeter (1943) have argued that a certain degree of profitability derived from the use of market power is necessary to achieve optimal innovation through research and development expenditures. From this perspective it is interesting to note that the two large private French water companies maintain large research budgets, and are recognized world leaders in developing innovative technology in the water and wastewater industry.

7.2.2 Transaction Costs and Contracting⁷⁸

Transaction costs refer to the information costs of negotiating and managing contracts, or the costs of making a market. Examples of transaction costs are bargaining, enforcement, legal and administrative costs required when contracting exchanges of goods and services. The concept was originally developed by (Coase, 1946) to explain the development of firms, which Coase posits exist precisely to minimize transaction costs by moving to a different organizational model than the market, the firm with its hierarchy of command.

Transaction cost economics, as further developed by (Williamson, 1976, 1985) derives from two basic assumptions about the behaviour of economic agents (e.g. firms, individuals), bounded rationality and opportunism. Bounded rationality refers to the inability of firms and individuals to possess full information regarding the behaviour and intent of others, while opportunism assumes that agents will take advantage of these information asymmetries to benefit themselves.

⁷⁸ The section draws heavily on (Dahl, 1998).

Contractual arrangements are linked strongly to the nature of institutions. Analysts have used three sets of factors relating to institutions to explain the nature of contractual arrangements: uncertainty, asset specificity, and the frequency of transactions.

Uncertainty refers to the possibility for unpredictable and complex future states of the world, where given bounded rationality, increases in uncertainty will raise bargaining or noncompliance transaction costs as well as serve to generate a preference for a more formal relationship between agents. For example, high levels of uncertainty may generate lower transaction costs through a long-term contract, rather than short-term spot markets.

Asset specificity refers to the investment in relationship specific assets, such as capital (physical, human) or location. Asset specificity implies that many capital assets have a low value in the next best use, and are usually locked in, or sunk, in time. The larger the difference between the next best value and the historical value, the greater the possibility for opportunism or strategic behaviour. As assets become more specific to the transaction, the costs associated with monitoring and enforcement rise, and again transaction costs may be minimized through more formal long-term relationships.

Frequency of transactions refers to the number of individual possible market transactions between agents that could possibly take place. As frequency increases, the total transaction costs associated with each individual time period can be expected to rise, again possibly generating minimized transaction costs through a more formal long-term relationship. One example could be using a firm rather than individual labour sub-contracts.

The characteristics of bounded rationality, opportunism, uncertainty, asset specificity and frequency generate "the world of governance" (Williamson, 1985). This is characterized by three principal forms of governance; market, bilateral (e.g. long term contracts) and unified (firms; public ownership, public utilities). Water utilities are perhaps the most asset specific of all industries, and have a high frequency of use in the output market. In the output market this leads to long-term implicit, or negotiated, contracts between residential, commercial and industrial users. In terms of input markets the choice of governance is between long-term contracts, and unified forms such as public ownership or public utilities. The transaction costs involved in contracting are the cost of negotiating, managing, monitoring and enforcing the contract, while the transaction costs involved in public utilities are the costs of regulation. As (Goldberg, 1976) pointed out the design of a regulatory system can be viewed as an implicit "contract", with considerations similar to the drafting of a long term contract between private parties, where both opportunism and asymmetric information need to be addressed by the regulator. Public ownership minimizes input transaction costs by internalizing all costs related to market exchanges (e.g. contract, stock market) within the firm.

A contract franchise by a formerly publicly owned utility increases the transaction costs of utility management both in terms of letting, managing, and enforcing the contract. The high degree of asset specificity of water utilities leads to increased possibilities for opportunistic behaviour, both on the part of the bidder by, for example, renegeing on contract obligations, and the contracting authority, renegeing on payment. The success of contract bidding for water utilities in lowering costs relative to public ownership, therefore, depends to a large extent on both the good faith and contract negotiation and maintenance skills of the two parties involved.

The privatization of a formerly publicly owned utility increases the social transaction costs associated with utility ownership because of the invariable requirement for economic and water quality regulation. The success of public utility governance also depends to a large extent on the good faith of both parties, and the regulator abilities exhibited by the regulatory agency.

7.2.3 Incentive Regulation⁷⁹

Several major advances in the theory of utility regulation can be grouped under the heading of incentive regulation. Two complimentary developments are price cap regulation and bench marking. These developments grew out of the realization that the traditional form of utility regulation, rate of return regulation (RORR), distorted the incentives to the utility to cost minimize over time. RORR refers to a process of periodic rate setting reviews by a regulatory agency where the firms costs are established, and prices set to equal total accounting costs, plus some form of acceptable return to shareholders, normally a return on capital equivalent to that prevailing in other sectors of the economy. Given this regulatory structure, utilities have incentives to both distort their investment plans, and provide misleading information to regulators in order to maximize permitted profits. Some of the distortions noted in economic theory are;

- a tendency to over capitalize (Averch and Johnson, 1962);
- a tendency towards high risk oversized, capital projects (Lyon, 1991);
- inefficient decisions regarding multiple service offerings (Breutigam and Panzar, 1989).

Incentive regulations seeks to both de-couple the relationship between costs and revenues, and provide incentives for utilities to provide accurate information over time, thus leading to optimal planning and output decisions. Two principal practical forms of incentive regulation of relevance to water utilities are price cap regulation and benchmarking. Price cap regulation, as proposed by (Littlechild, 1983) is a deceptively simple idea: set a fixed price ceiling on the price (or revenues, given multiple services) the utility can charge, and the utility will have the same incentives to minimize costs and maximize profits as in competitive markets. Given uncertainty about future conditions there is a need to allow flexibility, leading to an index approach to changing prices in the short run. The classic index advanced was $RPI - X$, where RPI refers to retail price index (or consumer price index in Canada) and X refers to the expected productivity improvement. The third necessary component is periodic formal price reviews, where the previous efficiency gains realized by utilities are captured in the form of price or revenue cap reductions, as well as through recalibrating the index (i.e. determining X) for the next period.

The necessity for periodic reviews, as well as the initial setting of the price cap, requires elements of RORR to creep in. Most notably in determining the utilities costs to setting the initial price cap, and then in assessing future capital investments in each subsequent price review. The timing of periodic price reviews is also important as utilities must have sufficient time to make and retain profits through cost efficiencies, while not be too long in order to allow the benefits of reduced costs to be passed on to consumers.

An important supporting element in setting price caps, from the point of view of asymmetric information, is benchmarking or yardstick competition. Using yardstick competition, firms are rewarded, or punished, (i.e. their revenue caps are set) based on their relative performance compared to a similar set of utilities. Thus the utilities revenues are entirely divorced from their own cost structure, while shocks that effect all firms (e.g. energy price increases) will automatically be factored into the caps for all. The problem with yardstick competition lies in accurately placing utilities within homogenous groupings, given different possible cost conditions for water utilities.

The combination of price-cap and yardstick competition allows for two sources of relatively clear and unbiased information on utility costs to develop over time; a) historical data through the dynamic accounting records of specific public utilities at each periodic review, and b) cross-sectional data from different utilities at each review. Thus, in theory, price caps should naturally converge to relatively efficient levels over time, minus some degree of excess profit due to cost minimizing innovation in the period between each price review. The privatized water utilities in

⁷⁹ This section draws heavily on (Lyon, 1999).

England and Wales are regulated using price caps and benchmark competition; this regulatory form will be described in the next section on empirical results.

7.3 Empirical Studies Related to Utility Organization

While economic theory can provide the basis for making hypotheses about the impact of different forms of utility regulation, only empirical assessment can test those hypotheses. Our review of the empirical literature found that all empirical studies in this decade were in the countries of England, France, and the United States, with each country forming a distinct natural experiment in managing water utilities. The structure of this section will be based on each individual country.

7.3.1 France

France is the country with the longest and most comprehensive experience of franchise bidding for water utilities, where approximately 78 % of the population is provided service by private companies (Cameron, 2002). Municipalities (communes) that own the utility infrastructure are coordinated and partially supervised by river basin planning authorities, which both levy effluent charges, and distribute subsidies within the basin. The six river basin authorities report in turn to the water office of the Environment Ministry, which also provides some subsidies to the municipalities, from general tax revenues. In the French system mayors of the municipalities are both responsible and personally liable for the quality of water supply and sewage services.

The derivation and classification of the many possible forms of contracting for water services⁸⁰ derives primarily from the French experience where the three principal forms are;

- public operation, (regie direct);
- leasing (affermage), and;
- concession.

The different forms of organization differ in the scope in the scope of service provided by the contracted firm, from no private participation (public operation), to operation and maintenance of the system, including billing (lease), to full investment, operation and billing (concession). However in all cases the municipality (commune) continues to own the system. The case of management contracts (gestion intermediaire) where the private firm is not responsible for billing is not common in France, nor is full privatization. Table 7.1 lists the actual forms of contract arrangement used in France as of 1995, based on a sample of 2190 municipalities containing over 5,000 population.

Table 7.1 French Contractual Arrangements	
Type of Contractual Arrangement	Percentage
Public Operation	24
Management Contracts	2
Lease	57
Concession	5
Privatization	1
Others	2
Source: (Menard and Saucier, 2000)	

Menard and Saucier (2000) and Clark (2002) both conducted empirical studies that attempted to apply transaction cost economics to the French experience. The Menard and Saucier study used proxy data to test three propositions derived from transaction cost economics:

⁸⁰ See Cameron, 2002 or Menard and Saucier, 2000 for a full listing.

- the more a geographic area requires specific investments, the smaller the probability of outsourcing;
- the higher the uncertainty associated with investments meeting quality standards, the smaller the probability of outsourcing;
- local authorities with limited budgets are more likely to out source than provide the service themselves, where significant specific investments are required.

The study also tested for the relative performance of different models of governance.

This study was based on 2,190 utilities serving all French municipalities over 5,000 population for the period 1993 to 1995, approximately 73 % of the French population, and used logit regression methods to analyze the choice of governance. Proxies were used to represent investment needs; these proxies consisted of raw water quality, raw water source, and population size. Uncertainty was proxied by the river basin location, while limited budgets were proxied by the percentage of the permanent as opposed to seasonal population. Performance was assessed by whether or not water quality meets drinking water standards⁸¹. Reliable financial data on French water utilities did not appear to have been readily available, though both papers cited here indicated that efforts were underway to collect it.

The results relative to the mode of governance indicated that:

- the choice of governance is highly associated with the river basin location, where local authorities in some basins (e.g. Seine – Normandie) tend to delegate more, while others (e.g. Rhine- Meuse) delegate less
- large municipalities are much more likely to sub-contract out, and to prefer the concession mode;
- municipalities with poor raw water quality or ground water sources are more likely to choose public operation; and;
- municipalities with highly variable populations will tend to sub-contract more.

The authors summarized their results relative to transaction costs by stating that,

“The choice of a mode of governance proceeds in two steps. The decision to outsource depends crucially on the financial constraint, particularly when investments are major ones. If the decision is to outsource the choice between lease and concession depends largely on the density of the population and the concomitant investment ... where local authorities have much more control over the private operator with a lease”.

The results relative to performance, using potable water quality as an indicator of performance, indicated that, for comparably-sized municipalities, and approximately comparable levels of raw water quality, no statistical impact relative to the mode of governance was apparent.

In the second empirical study of French water systems, Clark (2000) assessed the impact of the personal liability, both criminal and civil, imposed on municipal mayors in France, and the prohibition in French law from ensuring themselves against this negligence risk. The study postulated that the law biases the choice of governance towards private contracts, as these contracts allow for liability to be passed from the mayor to the private company, thus benefiting the mayor, but not necessarily the municipality.

⁸¹ The question of the quality of drinking water in France is an interesting area for further investigation both the Menard study and the study cited in Cameron (2002) seem to indicate that drinking water quality may not meet WHO or EU standards for considerable periods of time. For example “ in most of Brittany ... nitrate levels exceed WHO limits, as a result tap water is unpotable for much of the year and in most places.

The study empirically determined the size of the insurance premium that would be required if permitted by law, and contrasts this cost with the additional costs imposed by private, rather than public, suppliers of water services. The study was based on a case study of a small (2000 inhabitants) municipality in the south of France, with a historical record of legal damages related to pollution incidents as imposed by the courts (cour des comptes). Clark concluded that

“(The study) address(ed) the agency conflict that pits mayor against commune in the choice of water management ... the economic cost of this conflict... can be modeled as the value of an insurance policy that pays all losses arising from accidents for which the mayor might be held responsible. ... In the case study the 3 % annual cost of the mayors personal liability contrasts with the 20 % or 30% jump in rates reported by Nowak (1995) when delegation is substituted by direct management.”

7.3.2 England and Wales ⁸²

In England and Wales, 10 privatized water utilities provide water and wastewater services, with the utilities organized as monopolies in each river basin. There are also 16 smaller private water supply companies operating within the regulated utilities, many of which date from the 1800s. The 10 utilities were privatized in 1989, as a replacement for a system of integrated river basin management (IRBM), while the water supply companies had been private previously. Three regulatory agencies were set up to monitor the industry, water quality, effluent and economic regulation, the latter of which is called OFWAT. The utilities were privatized as part of a wave of utility privatization in the United Kingdom, corresponding to sharp ideological return to a more neoclassical version of the role of the state, often referred to as “Thatcherism”. They were also under capitalized, and facing new requirements for improved water and sewage quality based on European directives. Finance was equity based through property taxes, and residential metering was rare. The utilities were sold for \$ 5.4 billion pounds, for a net loss of \$ 1.4 billion pounds, given debt forgiveness and initial subsidies. The initial offering was over subscribed, and several buy-outs by foreign firms, notably the French firms took place, as well as some mergers with other utilities.

OFWAT regulates the industry through a revenue-cap and yardstick competition approach, with rate reviews every five years. The initial rate setting period was characterized by increases in prices, profits, metering, potable water quality, sewage quality, investment and cut-offs. The industry recovered its initial investment within the first rate setting period through dividends. The combination of high profitability, increasing cut-offs (from 940 in 1989 to over 22, 000 at a peak in 1992), some service failures, as well as evidence of health impacts such as increased dysentery and even cholera, in lower income segments, generated considerable social unease with a privatization decision that had never been popular (Cameron, 2002).

However, the second rate setting period appears to have been calmer as OFWAT became more conscious of the nature of the required regulation, and the industry became more conscious of their social obligations. In setting the second rate-cap a fairly exhaustive examination of the utilities finances were conducted, including the requirement for detailed 20-year capital plans, not unlike what would take place under RORR. However, despite flatter permitted increases profitability remained quite high, and the 1999 revenue cap actually reduces prices in most utilities, by an average of 21 % annually, over the five year period, with however 10 to 14 % price reductions in the first year. This reflects increasing scepticism on the part of OFWAT regarding utilities reported required future capital expenditures. OFWAT has also been given new regulatory powers over services, and the utilities power in terms of cut-offs been reduced, as a result of the 1999 Water Industry Act regulations. (OFWAT, 2000).

⁸² This section draws heavily on (Cameron, 2002).

There are several empirical studies which attempt to assess the impact of privatization on the efficiency of the water industry, most of which focus on the initial 1994 price setting review process. Some results of two of the studies regarding economies of scale and scope in English water supply and sewerage can be found in chapter three. Possibly the most thorough study was by (Saal and Parker, 2000). This study assessed the performance of water utilities, both pre- and post-privatization, using utility data obtained from the utilities sale prospectus' as well as standard utility accounting in reporting to OFWAT, thus generating a complete data set from 1985 to 1999. The study used a trans-log cost function comprising water and sewer outputs, and input prices for capital, labour, and other operating expenses using the Zellner method of OLS, as per (Kim and Clark, 1988) using both quality adjusted and unadjusted models. The study examined two premises related to utility organization:

- the privatization of the water industry in England led to lower costs of production;
- regulatory tightening in 1994 led to efficiency gains.

The results indicated that a statistically significant reduction in the trend growth of costs did not occur after privatization. However, results do indicate that a statistically significant reduction in costs did occur after the 1994 rate tightening review. The authors suggested that, "The efficiency gains which occurred after privatization...are not attributable to privatization per se, but rather to the system of economic regulation that was implemented at privatization and made more stringent in 1994".

There are also several studies relating stock market returns to regulatory behaviour by OFWAT, most involving John Sawkins (e.g., Sawkins, 1996; Marona, 2000). The Marona study examined stock market volatility, under the premise that the 1994 review would have reduced volatility as industry became more predictably regulated. The 1994 review was generally considered to have been thoroughly done by OFWAT, even "arguably an object lesson in the UK regulatory process" as contrast to the repeated revisions to caps in the electric industry. The authors find that volatility was reduced, supporting the contention of a well-conducted review.

7.3.3 United States

The majority of American water utilities are municipally owned and operated, however anywhere from 15- 20 % of the population are served by investor owned utilities, many of which date from the late 1800s. These private water utilities tend to be small. Those that are large (e.g. over 1 million people) are formed by aggregations of smaller systems have been bought by one company. The three largest U.S. companies (American Water Works, United Water Services, Philadelphia Suburban) are organized as separate companies by state. Historically, the ownership is varied and changing with owners being institutions and individual stockholders, but there have been recent acquisitions by French and English water companies. Consolidation also occurs as companies buy each other out in order to consolidate geographic areas, or to expand their portfolios World Bank (1994). U.S. utilities typically report to a State Public Utility Board or Commission, which tend to be conservative and, in general, to rely on RORR. A small but growing trend is contracting out various water services, using the various French models. An example described in (Cameron, 2002) is that of the city of Atlanta.

There are several different empirical studies in the United States examining the question of public and private utility ownership undertaken in this decade, all involving Raffiee (e.g., Raffiee et al., 1993, and Bhattacharyya et al., 1994, 1995). This set of papers generally duplicated the results of previous studies (see Teeple and Glyer, 1988a, b), indicating a rather tenuous relationship between the form of ownership and cost efficiency, possibly supporting the growing economic consensus that the prime determinant of cost efficiency improvements in utilities is the degree of competition which can be introduced into the input and output markets, rather than the type of ownership. A good summary of much of the literature in this area can be found in Saal and Parker (2000).

The initial study in the United States (Raffiee et al, 1993) used 1989 survey data for the American Water Works Association (AWWA) of 238 public and 33 private water utilities. The econometric methods were based on the use logit OLS using the theoretical basis of the weak axiom of cost minimization to construct an efficiency index to compare the observed and optimum cost of production for each public and private water utility, a form of statistical benchmarking. The study concluded that the average costs are lower for private firms than for public firms, and that there is considerable deviation from minimum costs in both types of water utilities.

A second study (Bhattacharyya et al., 1994) used a 1992 AWWA survey of 225 public and 33 private water utilities. Again, using logit OLS, the cost function was estimated as a “generalized variable cost function that exhibits the regular characteristics of the neoclassical cost function but does not require that cost minimization subject to market forces be imposed as a maintained hypothesis”. The cost function assumed that capital is fixed in the short run, and that utility managers may have asymmetric information on input prices not observable to the economist. The results indicated that public utilities were more efficient than private utilities, but again with considerable deviation from minimum cost for both types of utilities. The study also found some evidence of over capitalization in both public and private utilities.

The third study (Bhattacharyya et al., 1995) used the same AWWA survey, but with 190 public and 31 private water utilities. A stochastic cost frontier is used to specify the cost of inefficiency in terms of different ownership structure and firm specific characteristics (e.g. length of pipe in distribution system, number of emergency breakdowns, service mix), using a translog production function, with a two stage logit OLS. The results indicated that both groups of utilities are cost - inefficient, though private firms are in general more cost-inefficient than public firms. However, smaller private firms were less inefficient than public firms in smaller municipalities, while public utilities were far more efficient than private utilities at higher levels of production. Similar to the results indicated in Chapter Three, the study also indicated that utility costs can be controlled by reducing the number of emergency breakdowns in all utilities, and that in large utilities cost can be minimized by controlling the expansion of the service area.

7.4 Innovative Financing Measures

This section reviews the theory and practice associated with some innovative financing mechanisms by which senior levels of government can subsidize municipal water and wastewater operations. It focuses on the experience of two common efficient and innovative financing measures; a) an effluent charge\subsidy measure (France) and b) revolving fund loans (United States). In England, with full privatization of water utilities, there are no subsidies to water utilities, with all capital originating in private markets. These measures contrast with the standard subsidy measure of grants or loans derived from general tax revenues, as were common in Ontario until relatively recently.

7.4.1 Effluent Charges and Subsidies

As is discussed more extensively in Chapter Three, in economic theory the effluent charge is considered to be an efficient method of regulating externalities associated with water pollution. The main rationale is the charge allows polluters the choice of the effluent reducing technology or practices required to attain social abatement targets, therefore being preferable to mandatory standards. A side benefit of effluent charges involves the revenues raised, which can then be used to subsidize other social projects, such as improving municipal water and wastewater technology, or financing water quality enforcement and management administration. The revenues raised from effluent charges are efficient compared with other forms of taxation, because they do not impose constraints on social goods such as effort (income taxes) or profits (corporate taxes), but derive

from the regulation of bads (e.g. pollution). Effluent charges are quite common in practice, including Canadian provinces, notably British Columbia, many U.S. states, and several European countries (e.g. France, Germany, Holland). In North America the charges are generally lower, funding primarily water resource administration. In Europe the charges also support subsidies for municipal utility and other wastewater treatment operations.

In France, the program for water management is primarily run by six river basin authorities, which have levied effluent taxes since 1968, as autonomous self-funded bodies of the French government. The charges are on a range of different emissions, including suspended solids, oxidisable matter, nitrate levels and various chemicals and heavy metals. The charges fund administration and subsidies, with subsidies also available from the Environment Ministry. There are also tax advantages available to utilities, including high depreciation allowances, refunding of sales revenue to public utilities, and reduction of asset values for property tax purposes (Bongaerts and Kraemer, 1989). A 1989 OECD report concludes that the charges were too low to have a pronounced positive impact on effluent quality, and were used primarily as a revenue raising mechanism. However, charges have risen steadily in the 1990s, with total revenues from charges of approximately \$ 1.6 billion U.S. as of 1996 (OECD, 1997).

7.4.2 Revolving Funds

Revolving fund finance is an alternative to the more traditional subsidy mechanism of direct grants. Revenues from senior levels of government are used to establish a revolving fund out of which loans are made at below market interest rates. The money is then repaid into the fund and can thus be used again as a continuous source of funding. The possible efficiency advantages from revolving funds stem from the degree of risk assumed by the senior levels of government, which pay lower interest rates on borrowing than do municipalities, and where some smaller municipalities may have trouble financing investment. This stems from the less risky nature of senior levels of government debt. The advantages from a senior government perspective are that funds allow for continuing involvement in utility management, possibly due to equity considerations, but at a lower price than grants. Government subsidization is generally regarded as inefficient in economic theory because it distorts market prices, leading to underpricing of services, and excessive use. Areas where subsidization is less inefficient include areas where negative externalities are present, such as effluent treatment, metering, and water quality improvements. Equity considerations are also often justifications for subsidization, where a relatively common consideration is the higher average costs of small community systems.

Prior to 1987, the U.S. federal government provided direct grants to municipalities for water treatment through a program of construction grants to localities under the Clean Water Act, including a national target of universal secondary level treatment. As of the Water Quality Act of 1987 the grant program was ended, and the state revolving fund (SRF) program instituted. The SRF is run through the federal Environmental Protection Agency (EPA) and allocates money to different forms of state organizations, which must contribute a minimum of 20 % of matching state funds. SRF funds are then loaned to municipalities within the state at advantageous rates. There is considerable liberty granted in the use of federal funds, including use as collateral to leverage additional borrowing for the state fund, or to issue state fund bonds. The ongoing costs to the federal government are regular top up costs to account for defaults and low interest rates. The SRF program funds both wastewater and water quality improvements, with annual federal contributions being approximately \$ 1.3 billion for wastewater, and \$ 825 million for drinking water as of 2001. (Bonds, 2001) As of 1995 total capitalization of SRFs was \$ 16 billion, of which the federal share was 11.1 billion. (EPA, 1995). An additional tax advantage in the United States is tax-free municipal bonds.

7.5 Summary

This chapter has reviewed some of the economic theory and empirical results related to network utility management and organization. The principal areas addressed in terms of theory include; a) the opening of markets to competition, b) transaction cost economics and c) incentive regulation. The two main areas for introducing competition pressures to water utilities are the contracting out of input services through competitive tendering, and privatization combined with competition imitating incentive regulation (e.g. price and revenue caps, benchmarking). Transaction costs refer to the costs of generating markets, with transaction costs relative to inputs being the costs of tendering, managing and enforcing contracts. With the privatization option, transaction costs refer to the costs of health, environmental and economic regulation. The possible efficiency gains from moving to either option from public ownership and operation will depend on the trade-off between increased transaction costs and the potential cost efficiencies generated by increased competition.

Empirical studies correspond to three countries, each constituting a natural experiment in water utility management (France, England, United States). France is the country with the longest, and most extensive, experience with contracted utility services, with over 70 % of utilities contracting all or some of their services to private firms. However, financial information on French utilities does not appear to exist. An empirical study identifies the main trigger for moving from public to private operations as being the financial constraints faced by the municipality. Comparisons of utility performance based on water quality indicate that both public and private operators perform equally well, controlling for the size of municipality and raw water quality. In France publicly operated utilities tend to be small, and have poor raw water quality. A second study suggests that the use of private contracts in France are higher than optimal due to a quirk in French law whereby municipal mayors are personally liable for water quality problems, and by law, cannot insure themselves against this liability. They can however, transfer this liability to private firms through outsourcing.

England is the only country in the world with full privatization of water utilities, as of 1989, with 10 large private utilities based on river basins. The industry is regulated by agencies for water quality, pollution, and economic regulation (OFWAT). OFWAT regulates the industry through revenue caps and yardstick competition, with cap reviews every five years. The initial rate setting period was characterized by an abrupt increase in prices, profits, metering, potable water quality, sewage quality, investment and cut-offs. The combination of high profitability, increasing cut-offs, and some service failures, generated considerable social unease with a privatization decision that had never been popular. However, the second rate setting period appears to have been calmer as OFWAT become more conscious of the nature of the required regulation, and the industry became more conscious of their social obligations. Price caps have declined over each rate review as profitability has remained high, and the 1999 revenue cap actually reduces prices in most utilities, by an average of 2.1 % annually, over the next five-year period. with 10 to 14 % price reductions in the first year. OFWAT has also been given new regulatory powers over services, and the utilities power in terms of cut-offs been reduced, as a result of the 1999 Water Industry Act. Empirical studies of the English experience have found cost efficiency improvements related not from privatization per se, but to privatization and the system of economic regulation instituted by OFWAT.

In the United States most utilities are publicly owned and operated, however anywhere from 15 to 20 % of the population is served by privately owned utilities, regulated under rate of return regulation by state public utility commissions. Empirical studies in the United States find a tenuous relationship between ownership and efficiency, with public utilities appearing to be more efficient than private utilities in general. However, private utilities appear to be more efficient in smaller municipalities.

Two innovative financing mechanisms described are an effluent charge and a revolving fund. An effluent charge is a fee placed on loadings of pollutants, with levels generally set to cover administrative costs of water management, as well as possibly subsidize water and wastewater

treatment. It is used in France as a source of funding by six self-funded river basin agencies reporting to the ministry of the environment. Total annual revenues as of 1996 were approximately \$ 1.6 billion U.S.

A revolving fund is a dedicated fund which loans money to municipalities at below market interest rates. The United States uses a state revolving fund administered by the federal Environmental Protection Agency. All contributions from the federal government require a 20 % state matching contribution, and the federal monies can be use to leverage other finance, for example as collateral for fund bonds. Annual federal contributions in 2000 were \$ 1.3 billion for wastewater and \$ 825 million for water, with a capitalization of approximately \$ 16 billion.

CHAPTER 8: PRINCIPAL FINDINGS

This report has presented a detailed analysis of the economic aspects of municipal water servicing. It has described, analyzed, and synthesized the economic principles and considerations that pertain to water servicing infrastructure. These principles were used to analyze three key areas. The first of these was the pricing of water services. Here, the focus was on full cost recovery and economic efficiency. The second was an examination of the economic characteristics and implications of alternative organization, ownership and management arrangements for municipal water utilities. The third area of emphasis was the long-term management and capital financing of water utility assets.

Before addressing these three issues, the report: 1) outlined the principles currently in use for assuring adequate financing and long-term system sustainability; 2) placed water utilities into the broader context of water resource economics; and 3) looked in detail at the characteristics of demand and supply for municipal water services.

Although geographically, the focus of the report is the Province of Ontario, the Ontario is far from alone in facing economic and financial problems with its municipal water utilities. On the contrary, these problems are occurring in many parts of the world – hence, the importance of examining the research and experiences as widely as possible.

8.1 Introduction

The major problem facing municipal water utilities in Ontario currently can be summarized succinctly: there is a lack of monetary resources being devoted to the municipal water utility “industry”, particularly given increasing demands on public funds, the transfer of ownership of capital assets from the Ontario Clean Water Agency (OCWA) to the municipalities, and projected increased future spending requirements, as implied, for example, by new drinking water guidelines. These issues have major implications for the long-term physical and financial viability of Ontario’s municipal water industry. Some of the characteristics of this problem include: 1) insufficient financing of system capital; 2) inefficient rate setting methods; and 3) a possible under-utilization of resources available in the private sector as both utility operators and potential sources of capital.

The problem of inadequate capital investment funding is a long-term problem that may not be evident for many years, due to the long-lived nature of the assets involved in the water industry. However, the problem, once it becomes apparent, may be costly to remedy, and may have substantial public health implications. In this light it is interesting to note that capital spending by water utilities in Ontario was lower in 1998 and 1999, than in all of the previous eight years.

This report addresses each of these issues in an examination of possible alternatives to assure and improve the economic and financial viability of Ontario’s municipal water industry.

8.2 Key Principles and Economic Concepts

Chapter Two provided an overview of the water utility industry in Canada, which, for the most part, has been publicly owned and operated. For this reason, several over-arching principles, not all of them economic, have been used from place-to-place and time-to-time to govern operations, including:

- economic efficiency;
- full cost recovery;
- enhanced market competitiveness;
- equity;
- practicality; and
- environmental sustainability.

These principles may conflict with one another, sometimes explicitly, but more often implicitly. The chapter cites several such conflicts. Also, from an economic viewpoint, the most analytically viable of these principles is the first one – economic efficiency – a concept that underlies much of this report. It should also be noted that the use of multiple principles for operating water utilities stands in marked contrast to the main guiding principle used in the private sector – namely profit maximization. Interestingly, the latter is based on the principle of economic efficiency, which is the central one advanced in this report as the economic cornerstone for the water utility industry. In later sections of the report we review some empirical studies assessing increased private sector involvement in the provision of water utility services.

Following from these principles, a number of basic economic concepts, which make the municipal water servicing industry quite challenging to deal with. These concepts include:

- marginal cost pricing;
- economies of scale and scope;
- natural monopoly;
- public goods;
- externalities;
- water as renewable, but depletable, resource; and
- property rights.

Each of these concepts forms part of the fabric within which the water servicing industry must operate. Taken together, they illustrate some of the complexities that must be addressed by decision-makers in this field.

8.3 Water Utilities in an Water Resource Economics Context

Chapter Three examined some of the broad issues involved in water management in an economic context. The focus here was two-fold: the impact of water management approaches on water utilities; and an identification of some of the economic issues involved and how they affect the water utilities. These utilities are both affected by, and affect, the quality and quantity of in-stream water, because they both withdraw in-stream water as a basic input in the production of potable water, and discharge effluent to receiving water bodies.

The economic concepts of public goods, property rights and externalities were shown to be central ones used in examining water resource management issues. The central economic problem here involves the efficient allocation of property rights to a publicly owned resource in order to address externalities associated with its use. Under the Canadian constitution provinces own the in-stream water, while the federal government has authority over fisheries, navigable water and international waters. These constitutional realities form a framework within which rights can be assigned. The use of economic concepts in attempting to achieve the allocation of “usufructory” (or user) rights has largely been overlooked in the past.

The basic forms of possible property rights examined in this report were allocation through legal “rights” (e.g., those obtained from regulations); pricing charges and subsidies; and tradeable permits. Though economists have concentrated on the efficiency value of charges and permits,

common practice favours the use of legal instruments, such as effluent regulation or water rights, and subsidies. Nevertheless, the use of effluent charges is common in Europe, especially in France, Germany, and the Netherlands. Tradable effluent permit systems have been piloted in the United States, while various forms of water rights trading are being piloted in Australia, and the South West United States. Thus, the relatively newly developed field of water resource economics provides several insights into the water management problems currently being experienced. The use of economic instruments, such as effluent charges or tradable effluent permits, may offer new alternatives, and, in particular, possible supports for future management.

In terms of water quantity, Ontario is relatively water rich, though local conditions associated with groundwater may generate occasional scarcity. Accordingly, the system for allocating property rights to water for intake is relatively simple, consisting of provincially administered water taking permits, and small administrative fee.

Water quality problems are more prevalent, and include both point and non-point sources of pollution. Common sources of water pollution that have impacts on water utilities include traditional pollutants (turbidity, suspended matter), agricultural run-off, and toxic substances. The relative extent and complexity of water pollution externalities gives rise to a more complex system of property rights allocation, including separate provincial and federal regulation of point source effluent, federal regulation of toxic substances, and joint provincial-federal cooperation through the international Great Lakes Water Quality Agreement. The use of substantial subsidies has also been historically associated with these programs, with subsidies declining in the recent past. Addressing non-point source pollution, such as agricultural run-off, is underway in Ontario, but is not as well advanced as studies in the point source area.

Municipal water utilities are regulated by a provincial permitting system termed "Certificates of Approval" (CAs), based on ambient water quality (Ontario Water Quality Objectives) and best available technology. To meet these objectives, there were 600 waste treatment plants, both municipal and industrial, discharging treated wastewater into Ontario watercourses in 1998. Of these, 243 municipal systems were operated by OCWA, 207 by individual municipalities, and 163 by individual industries. Of the latter, 152 discharged treated process wastewater. Municipal effluent accounted for approximately 55 % of total discharge of total effluent volume regulated, compared to 45 % for industry. BOD production associated with agricultural operations (e.g. manure) represented 5.4 times that of municipal waste, is untreated and unregulated, but is not discharged directly into receiving bodies. A recent report by the Province assessing compliance of municipal wastewater plants with their CAs indicated that 28 % of plants were in compliance, and that almost 60% of the plants had inadequate data to assess compliance with the CAs.

Theoretical econometric studies related to pollution indicate a relationship between raw water quality parameters, such as turbidity or variability of supply, and increased water supply treatment costs. Some cost estimates related to the costs of pollution to water utilities are presented, including economic costs of \$ 155 million from Walkerton, and \$ 125 million over ten years for groundwater contamination from chlorinated solvents. Possible upgrading costs related to water supply and effluent treatment are found in Chapter Six.

8.4 Characteristics of Demand for Water Services

Chapter Four focused on the demand for municipal water services. For the purposes of this report, water demand was used to refer to the demand for both water supplies and wastewater treatment. Records from the 1989 to 1999 period demonstrated that municipal demands have grown in Ontario, but still form a small proportion of total water demands. Municipal water demand in Ontario is divided approximately 60 % residential, and 20 % commercial and industrial, respectively.

Municipal water demands are not evenly spread through time, but show pronounced peaks on a daily and seasonal basis. These peaking characteristics form important considerations in setting water rates, particularly pricing based on predictable seasonal patterns of use, for example, summer peak load pricing.

Water is a “normal” good in the economic sense, in that as price rises, demand falls. A large number of empirical studies have confirmed this characteristic. This is a key finding for public policy. Demand responds to price changes in various ways, depending on the type of demand, the time of year, the short- versus the long-run, and other factors. This relationship is measured by the concept of price elasticity of demand, which generally reflects the availability of substitutes to the use under consideration. In general, the fewer the substitutes, the lower the price elasticity. Many empirical studies have permitted price elasticities to be defined within relatively narrow ranges. Generally, most municipal water demands are price inelastic, with the most inelastic values (i.e., the lowest) occurring for indoor demands (-0.2 to -0.4). Outdoor elasticities (e.g., lawn watering) are relatively higher (-0.4 to -0.6), with industrial elasticities often higher still.

Econometric studies in the municipal water demand field have tended to concentrate in the residential demand area, with much less attention paid to commercial, (publicly-supplied) industrial, and public demands. Prices of other inputs also have an impact on water demand, especially prices of waste treatment and energy.

Income levels have a positive impact on residential water demands, in that as average income levels rise, so also do water demands. Income elasticity values are in the range of (+) 0.4. Metering water demands can have a substantial impact on water demands. Complete metering in a municipality, combined with appropriate pricing, has been shown to lower demand levels by up to 50% over flat rate, unmetered uses.

Water demand management presents a philosophy of management that focuses on the demands made on water resource systems. Demand management can have significant impacts on system costs by fostering lower levels of demand, and thereby lower system capital and O&M costs

8.5 Characteristics of the Supply of Water Services

Chapter Five assessed some of the cost characteristics associated with the production of water utilities, using both engineering and economic methods to determine utility costs. In engineering cost studies, the accuracy of cost estimates depends on the amount of data available to characterize the site and the project requirements. At a master planning level, study or order of magnitude cost estimates are typically used. The anticipated accuracy associated with this type of estimate is plus 50% or minus 30%. Common methods used to generate estimates include standard cost factors and curves, rules of thumb, and computer software. In general main trunk components (e.g. plant, intake, storage) are sized based on maximum day demand, while the distribution network is sized based on area, density and service mix (i.e. industrial, residential).

Cost curves are based on in house or published data. Published data are typically derived from municipal surveys of costs, notably those conducted by the U.S. Environmental Protection Agency. It is important that sufficient information regarding the source of curves and other data be available to so that all appropriate factors can be applied to deal with other costs such as inflation, site location, currency (\$US or \$CDN), contingencies, etc. Clear information allows all costs to be included once, and not repeated.

Additional costs that must be included when preparing a cost estimate include engineering fees, administration and legal costs, land acquisition costs, and provincial and federal taxes. The allowance for these items is dependent on the size and complexity of the project.

Study or order of magnitude estimates are used to compare the economics of various treatment options or the costs of major process components. It is important to realize that this type of estimate does not represent the actual construction and operation and maintenance costs of the project. Actual project costs are site-specific, cannot be generalized, and must be developed based on specific project requirements.

Econometric studies cover the water supply and sewage costs of four principal variables or types of variables, plant capacity, the rate of capacity utilization, spatial variables (e.g. area, density), and the service mix (e.g. proportion of industrial, residential, and commercial connections).

Results indicate that given fixed intake supply and health standards, water utilities exhibit long run increasing returns to scale, long run economies in both capital and operating costs, as well as short run economies in operating costs as capacity utilization increases. The result for capacity utilization reflects the discrete nature of capital investments in plant capacity, where plants are normally built to service demand in any given region for a considerable period of time, possibly 20 to 25 years into the future. Thus at any given time the majority of plants in a cross-sectional analysis will be operating considerably below capacity, where as demand increases and more capacity is used, increasing returns result. In an individual utility increasing returns can be expected to convert to decreasing returns at or near full capacity utilization (as in standard micro-theory), at which point a new plant would be built for the next 20 to 25 years.

The spatial and service mix variables determine the returns to scale within the distribution system, where population density and area served by the water utility act to determine the length of pipe and energy costs needed to service water and sewage, while the service type influences pipe capacity per connection. These distribution costs demonstrate decreasing returns as the area supplied by a given utility increases, and as density declines towards the perimeter of the service area. As the service type determines pipe capacity an increasing proportion of residential connections can also be expected to produce decreasing returns in the distribution network.

This trade-off between increasing returns relative to production, and decreasing returns as the distribution network expands determine an optimal system size (or minimum efficient scale). In any given community the optimal size will be determined by the interrelationship between the variables outlined above. Economies of scope in supply appear to exist between services to distinct user classes (residential, industrial), while economies of scope between sewage and water supply are not found in the literature, possibly indicating some scope for disaggregation of typical water utilities. An Ontario study estimating water supply and sewer marginal costs found that marginal costs in Ontario far exceed prices charged.

Additional considerations in determining the cost function for any given utility are changes in technology, due to treatment method, or source of supply. These changes serve to generate increasing returns to scale, as lower cost sources would have initially been used. There are also possible multi-plant economies of scale through shared administration and expertise in either large urban areas, or large specialized corporations.

8.6 Water Utility Pricing Theory and Practice

Chapter Six addressed the economic theory and practices of water pricing both in general and in Ontario. Initially, the chapter discussed the concept of full cost pricing, suggesting three efficiency-based criteria for defining full cost recovery based on marginal cost pricing;

- peak load pricing;
- recovery of marginal distribution network costs through connection charges;

- forward looking volumetric rates, incorporating long run marginal cost pricing, and estimates of future, rather than sunk, capital costs, and;
- extra-strength sewer surcharges associated with industrial effluent.

The chapter then addressed some of the economic solutions to the problem of full cost recovery in utilities with large economies of scale, such as water utilities, where average costs exceed marginal costs throughout most of the range of water demand. This problem has resulted in several different suggestions as to how to achieve both prices based on marginal costs and full cost recovery. The main solutions to this revenue gap were proposed in the chronological order in which they were published:

- subsidization by senior levels of government, and pricing at marginal costs
- Coase two part tariff: Volumetric pricing at marginal cost and imposing an additional fixed access fee on public utility users
- Ramsey pricing: Mark-ups on marginal costs based on charging users with inelastic demands higher prices than users with elastic demands
- Pareto-optimal linear outlay schedules, or volume discounts for large users

Some discussion was also devoted to increasing block rates, which are methods for ensuring full-cost recovery in utilities generating excess profits due to diseconomies of scale, for example drought in water scarce areas. Increasing block rates can also function as a means of transferring revenues from large water users to smaller water users on equity grounds, and is often recommended on water conservation grounds.

In addition to these proposed rate systems, economic theory has developed one other major theoretical advanced based on efficiency maximization;

- peak load pricing, or seasonal summer marginal cost pricing, in the case of water utilities

This implementation of peak load pricing was assessed under the assumptions of sub-additivity in the long run production function, indivisible capital, exogenously increasing demand, and optimal capacity sized to meet peak demands. The analysis indicated that under these assumptions, and using the standard definition of costs, there is no continuous long run marginal cost function (LRMC), simply a succession of discontinuous short run marginal cost (SRMC) functions corresponding to optimal discrete capacity increments. Thus the peak and off peak prices correspond to the intersection of the respective demand curves and short run marginal costs, and off-peak prices may exceed peak prices, depending on the particular time period within the capacity cycle. In general, peak and off peak prices will be less than average cost, necessitating an additional access charge to achieve full-cost recovery.

The chapter suggested a means of “rehabilitating” long run marginal cost as a means of setting municipal water rates and prices, in order to avoid the price instability associated with pricing at short run marginal cost. The section focuses on a discussion of various methods for estimating LRMC, given discontinuous capital. The concept of Turvey cost, and the extension to it, generate a continuous capital cost function which approximates the movement in SRMC over time, rising prior to capacity increments and falling afterwards, but smoothed to reflect all discounted future capital costs over the long run planning horizon. From this perspective LRMC is defined as a discounted moving average of future capital expenditures.

All of the rate structures derived from economic theory are based on maximizing social welfare, as discussed in the section of Chapter Two dealing with Pareto optimality. However, only two of these

structures are considered to be “first best” solutions, or solutions which allow for prices equal to marginal costs; these structures are (i) the Coase two part tariff, under certain conditions, and (ii) peak load pricing. The condition for Pareto optimal pricing occurs only if the access charge does not induce any users to disconnect or not connect at all, to the utility grid. A first best solution requires that the marginal cost equal price, while second best solutions minimize the social costs associated with divergence from the marginal cost-pricing rule, in a static setting.

One of the rare rate setting manuals for water utilities incorporates marginal cost pricing is the manual for the Canadian Water and Wastewater Association (CWWA, 1993). This manual uses both the Coase two-part tariff and peak load pricing as the basis for their rate setting methodology.

In contrast, the prevalent rate setting methods in use in North America, and Ontario, are engineering methods based on average cost pricing, with cost allocation to user groups based primarily on equity criteria. The expressed rationale is that, because residential users have more pronounced peaking behaviour, it is fair that they should be allocated a larger share of the sunk or historical capital costs, thereby generating the declining block rates commonly found in current rate setting practice. The two principal rate setting methods advocated by the AWWA are:

- the base extra capacity method, which allocates costs between base and “extra” capacity; and
- the commodity demand method, which allocates costs between commodity and capital costs..

These primarily engineering rate setting methods described above have the interesting property of approximating Ramsey pricing, in generating cost recovery by charging users with higher demand elasticities lower prices.

Finally the chapter assesses the degree of full costs recovery in current municipal water system operations are being recovered in Ontario and concludes that local municipalities are bearing most of the costs for water servicing. However, there is some debate as to whether the local, user-based funding system currently in place will ensure adequate economic and financial viability for water utilities in the future. Thus, even though full cost pricing seems to be in place, this may be illusory, because prices and the related level of investment appear to be below the level required to upgrade and maintain current water infrastructure. Some empirical estimates of this revenue inadequacy are provided, with annual estimates ranging from \$ 498 million to \$ 2.2 billion, depending on the type of upgrading specified.

8.7 Economic Theory and Empirical Results Related to Water Utility Organization and Management

Chapter Seven examines the economic theory and empirical results related to water utility management and organization, from the perspective of water utilities as network utilities. The principal areas addressed in terms of theory include; a) opening of markets to competition, b) transaction cost economics and c) incentive regulation. The two main areas for introducing competition pressures to water utilities are the contracting out of input services through competitive tendering, and privatization combined with competition imitating incentive regulation (e.g. price and revenue caps, benchmarking). Transaction costs refer to the costs of generating markets, with transaction costs relative to inputs being the costs of tendering, managing and enforcing contracts. With the privatization option, transaction costs refer to the costs of health, environmental and economic regulation. Possible efficiency gains from moving to either option from public ownership

and operation will depend on the trade-off between increased transaction costs and the potential cost efficiencies generated by increased competition.

The empirical studies included in the chapter focussed on three countries, each constituting a natural experiment in water utility management (France, England, United States). France is the country with the longest, and most extensive, experience with contracted utility services, with over 70 % of utilities contracting some or all of their services to private firms. However, financial information on French utilities does not appear to exist. An empirical study outlined in the chapter identified the main trigger for moving from public to private operations as being the financial constraints faced by the municipality. Comparisons of utility performance based on water quality indicate that both public and private operators perform equally well, controlling for the size of municipality and raw water quality. In France publicly operated utilities tend to be small, and have poor raw water quality. A second study suggested that the use of private contracts in France are higher than optimal due to laws making municipal mayors personally liable for water quality problems. By law, these officials cannot insure themselves against this liability; however, they can transfer this liability to private firms through outsourcing.

England is the only country in the world with full privatization of water utilities. As of 1989 10 large private utilities were based on river basins. The industry is regulated by agencies for water quality, pollution, and economic regulation, the latter of which is called OFWAT. OFWAT regulates the industry through revenue caps and yardstick competition, with cap reviews every five years. The initial rate setting period was characterized by an abrupt increase in prices, profits, metering, potable water quality, sewage quality, investment and cut-offs. The combination of high profitability, increasing cut-offs, and some service failures, generated considerable social unease with a privatization decision that had never been popular. However, the second rate setting period appears to have been calmer as OFWAT become more conscious of the nature of the required regulation, and the industry became more conscious of their social obligations. Price caps have declined over each rate review as profitability has remained high, and the 1999 revenue cap actually reduces prices in most utilities, by an average of 2.1 % annually, over the next five year period., with 10 to 14 % price reductions in the first year. OFWAT has also been given new regulatory powers over services, and the utilities power in terms of cut-offs been reduced, as a result of the 1999 Water Industry Act. Empirical studies of the English experience have found cost efficiency improvements related not from privatization per se, but to privatization and the system of economic regulation instituted by OFWAT.

In the United States most utilities are publicly owned and operated, however from 15 to 20 % of the population are served by privately owned utilities, regulated under rate of return regulation by state public utility commissions. Empirical studies in the United States have found a tenuous relationship between ownership and efficiency, with public utilities appearing to be more efficient than private ones in general. However, private utilities appear to be more efficient in smaller municipalities.

Two innovative financing mechanisms are an effluent charge and a revolving fund. An effluent charge is a fee placed on loadings of pollutants, with levels generally set to cover administrative costs of water management, as well as possibly subsidize water and wastewater treatment. It is used in France as a source of funding by six self-funded river basin agencies, which report to the Ministry of the Environment. Total annual revenues as of 1996 were approximately \$ 1.6 billion U.S.

A revolving fund is a dedicated fund which loans money to municipalities at below market interest rates. The United States uses a state revolving fund administered by the federal Environmental Protection Agency. All contributions from the federal government require a 20 % state matching contribution, and the federal monies can be use to leverage other finance, for example as collateral for fund bonds. Annual federal contributions in 2000 were \$ 1.3 billion for wastewater and \$ 825 million for water, with a capitalization of approximately \$ 16 billion.

REFERENCES CITED

Alber, R., Adams J., and Gould, P. (1971). Spatial Organization: The Geographer's View of the World. Englewood Cliffs: Prentice-Hall Inc.

Al-Qunaibet, M. and R. Johnston (1985). "Municipal Demand for Water in Kuwait: Methodological Issues and Empirical Estimates." Water Resources Research, 21(4): 433-438.

Altaf, M. and J. Hughes (1994). "Measuring the Demand for Improved Urban Sanitation Services: Results of a Contingent Valuation Study in Ouagadougou, Burkina Faso." Urban Studies. 31(10): 1763-76.

American Water Works Association Committee on Water Rates (1983). Water Rates 3rd Edition. New York: AWWA.

Averch, H. and L. Johnson (1962). "Behaviour of the Firm Under Regulatory Constraint." American Economic Review, 52: 1052-1069.

Bhattacharyya, A., E. Parker, and K. Raffiee (1994). "An Examination of the Effect of Ownership on the Relative Efficiency of Public and Private Water Utilities." Land Economics, 70(2): 197-209.

Bhattacharyya, A., T. Harris, R. Narayanan, and K. Raffiee (1995). "Specification of the Effect of Ownership on the Economic Efficiency of the Water Utilities." Regional Science and Urban Economics, 25: 759-784.

Bain, J., R. Caves, and J. Margolis. (1966). Northern California's Water Industry. Baltimore: Johns Hopkins University Press.

Baumann, D., J. Boland and W. Heinemann (1998). Urban Water Demand Management and Planning. Toronto: McGraw Hill.

Baumol, W., 1977, "On the Proper Test for Natural Monopoly in a Multiproduct Industry", American Economic Review, 67(5), pp. 809-822.

Baumol, W. and D. Bradford (1970). "Optimal Departures from Marginal Cost Pricing." American Economic Review, 60(3), 265-283.

Baumol, W. and W. Oates (1971). "The Use of Standards and Prices for Protection of the Environment." Swedish Journal of Economics, (March Issue), pp. 42-54.

Baumol, W. and W. Oates (1988). The Theory of Environmental Policy. New York: Cambridge University Press.

Beecher, J., P. Mann, and J. Landers (1992). Cost Allocation and Rate Design for Water Utilities. Columbus, Ohio: The National Regulatory Research Institute.

Billings, B. and D. Agthe (1980a) "Price Elasticities for Water: A Case of Increasing Block Rates" Land Economics 56(1): 73-84.

Billings, B. and D. Agthe (1980b) "Price Elasticities for Water: A Case of Increasing Block Rates: Reply." Land Economics, 57(2): 276-78.

Billings, R. and W. Day. (1989). Demand Management Factors in Residential Water Use: The Southern Arizona Experience. Journal, American Water Works Association, 87 (3): 584.

Billings, B. and C. V. Jones (1996). Forecasting Urban Water Demand. Denver: American Water Works Association.

Boggess, W., R. Lacewell and D. Zilberman (1993). "The Economics of Water Use in Agriculture" in G. Carlson, D. Zilberman and J. Miranowski, (eds.) Agricultural and Environmental Resource Economics. New York: Oxford University Press, 319-384.

Boisvert, R. and T. Schmit (1997) "Tradeoff Between Economies of Size in Treatment and Diseconomies of Distribution For Rural Water Systems", Agricultural and Resource Economics Review 27(2), 237-247.

Boiteux, M. (1949). "La tarification des demandes en point." Revue Generale de l'Electricite, reprinted as "Peak Load Pricing." Journal of Business. 30(2), pp. 157-179.

Boland, J., B. Dziegielewski, D. Baumann and E. Opitz (1984). Influence of Price and Rate Structure on Municipal and Industrial Water Use Fort Belvoir, Virginia: Engineer Institute for Water Resources, U.S. Army Corps of Engineers, Report 84-C-2,

Boland, John J., Alexander A. McPhail, and Eva M. Opitz. August 1990. Water Demand of Detached Single-Family Residences: Empirical Studies for the Metropolitan Water District of Southern California. Carbondale, IL: Planning and Management Consultants, Ltd.

Bonbright, J. (1961). "Fully Distributed Costs in Utility Rate Making." American Economic Review, 51(5), 305-312.

Bongaerts, J., and R.A. Kraemer (1989). "Permits and Effluent Charges in the Water Pollution Control Policies of France, West Germany, and the Netherlands." Environmental Monitoring and Assessment 12: 127-147.

Bower, B. (1966). "Industrial Water Demands." In Smith.S Water Research Baltimore, John Hopkins Press.

Braeutigam, R. and J. Panzar (1989). Diversification Under "Price-based" and "Cost-Based" Regulation." RAND Journal of Economics, 20: 373-391).

Bromley, D. (1991). Environment and Economy: Property Rights and Public Policy. Oxford: Blackwell and Company.

Brown and Caldwell, (1984). Residential Water Conservation Projects: Summary Report for U.S. Department of Housing and Urban Development. Walnut Creek, CA.

Brown, G. Jr., and R.W. Johnson (1984). "Pollution Control by Effluent Charges: It Works in the Federal Republic of Germany, Why Not in the U.S." Natural Resources Journal. 24: 929-966.

Brown, S, and D. Sibly (1986). The Theory of Public Utility Pricing. Cambridge, U.K.: Cambridge University Press.

Calloway, J., A. Schwartz and R. Thompson (1974). "Industrial Economic Model of Water Use and Waste Treatment for Ammonia." Water Resources Research 10(4): 650-8.

Cameron, T. and M. Wright (1990). "Determinants of Household Water Conservation Retrofit Activity: A Discrete Choice Model Using Survey Data." Water Resources Research. 26(2): 179-88.

Cameron, D. (2002). The Relationship Between Different Ownership and Management Regimes and Drinking Water Safety. Toronto: The Walkerton Inquiry, Discussion Paper.

Canadian Environmental Law Association (1998). "Ontario's Water Resources: The Need for Public Interest Regulation." by R. D. Lindgren and S. Miller. Toronto: Submission to the Ministry of Environment and Energy, Regarding the Water and Sewage Services Improvement Act, 1997.

Camp, R. (1978). The Inelastic Demand for Residential Water: New Findings. Journal. American Water Works Association. 70(8): 453-458.

Canadian Water and Wastewater Assoc., and Rawson Academy of Aquatic Science, (1993). Municipal Water and Wastewater Rate Manual. by J. Harris, D. Tate, M. Fortin, and M. Loudon. Ottawa-Hull: CWWA.

Carver, P. (1978). Price as a Water Utility Management Tool Under Stochastic Conditions. Ph.D. Diss. Department of Geography and Environmental Engineering, Johns Hopkins University Baltimore, MD.

Carver, P H. and J. J. Boland, (1980). "Short Run and Long Run Effects of Price on Municipal Water Use." Water Resources Research. 16(4): 609-616.

Cassuto, A. and S. Ryan, (1979). "Effect of Price on Residential Demand for Water Within an Agency." Water Resources Bulletin. 15(2): 345-353.

Chicoine, D., S. Deller and G. Ramamurthy (1986). "Water Demand Estimation Under Block Rate Pricing: A Simultaneous Equation Approach." Water Resources Research. 22(6): 859-863.

Clark, E. and G. Mondello (2000). "Resource Management and the Mayor's Guarantee in French Water Allocation." Environmental and Resource Economics. 15: 103-113.

Clark, R. and H. Goddard. (1977). Cost and Quality of Water Supply. Journal. American Water Works Association. 69(1): 13-15.

Clark, R. and R. Stevie (1981). "A Water Supply Cost Model Incorporating Spatial Variables." Land Economics. 57(1), pp. 18-32.

Coase R. (1946). "The Marginal Cost Controversy." Economica. 13 (8), 169-89.

Coase R. (1947). "The Economics of Uniform Pricing Systems." Manchester School of Economics and Social Studies. 15(5), 139-156.

Coase, R. (1960). "The Problem of Social Cost." Journal of Law and Economics. 3(10), 1-44.

Conley, B. (1967). "Price Elasticity of Demand for Water in Southern California." Annals of Regional Science. 1 (1): 180-189.

Crane, D. (1993). The Canadian Dictionary of Business and Economics. Toronto: Stoddart.

Crew, N. and P Kleindorfer (1979). Public Utility Economics. New York: St. Martins Press.

Dahl, C. and T. Matson (1998). "Evolution of the U.S. Natural Gas Industry in Response to Changes in Transaction Costs." Land Economics. 74 (3), 390-408.

Dales, J.H. (1968). Pollution, Property and Prices. Toronto: University of Toronto Press.

- Danielson, L. (1979). "An Analysis of Residential Demand for Water Using Micro Time-Series Data." Water Resources Research, 15(4): 763-767.
- Dearmont, D., B. Mccarl And D. Tolman (1998) "Costs Of Water Treatment Due to Diminished Water Quality: A Case Study In Texas" Water Resources Research 34(4): 849-853.
- Delcan (2001). Wastewater Collection and Treatment.. Toronto: The Walkerton Inquiry.
- Demsetz, H. (1967). "Toward a Theory of Property Rights". American Economic Review, 57: 347-359.
- Demsetz, H. (1968). "Why Regulate Utilities?" Journal Of Law and Economics. 11,(4), 55-65.
- Dinar, A. and A. Subramanian, eds. (1997) Water Pricing Experiences: An International Perspective. World Washington, D.C.: The World Bank, Technical Paper No. 386,.
- D. Dupont and S. Renzetti (1999) "An Assessment of the Impact of a Provincial Water Charge", Canadian Public Policy, 25(3): 361-378.
- Dupuit, Jules (1844) "On the Measurement of the Utility of Public Works", translated in International Economic Papers, 1952(2), 256-283.
- Dziegielewski, B. (1988). "Urban Nonresidential Water Use and Conservation", in Marvin Waterstone and R. John Burt (eds.) Proceedings of the Symposium on Water Use Data for Water Resources Management. American Water Resources Association, August, pp. 371-380.
- Dziegielewski, B. and E. Opitz, (1991). Municipal and Industrial Water use in the Metropolitan Water District Service Area, Interim Report No. 4 Prepared for the Metropolitan Water District of Southern California, Los Angeles. Planning and Management Consultants, Ltd., Carbondale, IL.
- Environment Canada,(1987). Federal Water Policy,Ottawa-Hull.
- Environment Canada (1995). Municipal Infrastructure Costs related to Capital PCE/TCE Groundwater contamination. Ottawa: Unpublished Report by Raven Beck Environmental Limited.
- Espey, M., J. Espey, and W.D. Shaw (1997). "Price Elasticity of Residential Demand for Water: A Meta-Analysis." Water Resources Research, 33: 1369-1374.
- FCM (Federation of Canadian Municipalities) (1985). Municipal Infrastructure in Canada: Physical Condition and Funding Adequacy. Ottawa-Hull.
- Feigenbaum, S. and R. Teeple (1983). "Public Versus Private Water Delivery: A Hedonic Cost Approach", Review of Economics and Statistics 65: 672-678.
- Feldstein, M. (1972). "Equity and Efficiency in Public Pricing." Quarterly Journal of Economics, 86: 175-187.
- Flack, J. E. (1965). Water Rights Transfers -An Engineering-ApproachStanford University. Palo Alto, CA. _ Ph.D. dissertation.
- Ford, J.and J. Warford (1969). "Cost Functions for the Water Industry." Journal of Industrial Economics. 18(11), 53-63.

Fortin M. and B. Mitchell (1990). Water and Wastewater Charges for Ontario: The User Pay Principle, Toronto, ON: Ontario Sewer and Watermain Contractors Association,

Fortin, M, E. Slack, M. Loudon, and H. Kitchen (2001). Financing Water Infrastructure. Toronto: The Walkerton Inquiry.

Foster, H. and B. Beattie, (1979). "Urban Residential Demand for Water in the United States." Land Economics, 55(1): 43-58.

Foster, H. S. and B. R. Beattie, (1981). "On the Specification of Price in Studies of Consumer Demand under Block Price Scheduling". Land Economics, 57(4): 624-629.

Fourt L. (1958). "Forecasting the Urban Residential Demand for Water," Agricultural Economics Seminar Paper, Department of Economics, University of Chicago; as quoted by Baumann et al, 1998.

Gallagher, D. and R. Robinson, (1977). Influence of Metering, Pricing Policy, and Incentives on Water Use Efficiency. Canberra, Australia: Australian Water Resources Council. Technical Paper No. 19.

Gardner, B. D. and S. H. Schick, (1964). Factors Affecting Consumption of Urban Household Water in Northern Utah. Logan, UT: Utah State University, Agricultural Experiment Station Bulletin. No. 449.

Gardner, R. (1977). "An Analysis of Residential Water Demand and Water Rates in Minnesota." Water Resources Research Center. Minneapolis: University of Minnesota, Bulletin 96.

Gibbs, K. (1978). "Price Variable in Residential Water Demand Models." Water Resources Research, 14(1): 15-18.

Golldberg, V (1976). "Regulation and Administered Contracts." Bell Journal of Economics, (7) 426-448.

Gottlieb, M. (1963). "Urban Domestic Demand for Water: A Kansas Case Study". Land Economics, 39(2): 204-210.

Griffin, R. and C. Chang. (1991). Seasonality in Community Water Demand. Western Journal of Agricultural Economics, 16(2): 207-217.

Grima, A. (1972). Residential Water Demand: Alternative Choices for Management. Toronto, Canada: University of Toronto Press.

Grunwald, O., C. Haan, D.. Debertin, and D. Carey. (1976). "Rural Residential Water Demand in Kentucky: A Econometric and Simulation Analysis." Water Resources Bulletin, 12(5): 951-961.

Hall, D. (2000). in Dinar, A. (2000) (ed.) The Political Economy of Water Pricing Reforms. London: Oxford University Press.

Halvorsen, R. and T. Smith (1986). "Substitution Possibilities for Unpriced Natural Resources: Restricted Cost Functions for the Canadian Metal Mining Industry" Review of Economics and Statistics, 68(3): 398-405.

Hanke S. and R. Wentworth (1981). "On The Marginal Cost of Wastewater Services." Land Economics 57(4): 558-567.

Hansen, L.G. (1996). "Water and Energy Price Impacts on Residential Water Demand in Copenhagen" Land Economics, 72(1): 66-79.

Hansen, R., and R. Narayanan. (1981). "A Monthly Time Series Model of Municipal Water Demand." Water Resources Bulletin, 17(4): 578-585.

Hardin, G. (1968). "The Tragedy of the Commons" Science: 162, 1243-1248.

Hartwick, J. and N. Oleweiller (1996). Natural Resource Economics. Toronto: Longham.

Hayek, F. (1945) "The Use of Knowledge in Society." The American Economic Review, 34(4), 521-530

Hayes, K. (1987). "Cost Structure of the Water Utility Industry" Applied Economics, 19: 417-425.

Headley, J. (1963). "The Relation of Family Income and Use of Water for Residential and Commercial Purposes in the San Francisco-Oakland Metropolitan Area". Land Economics, 39(4): 441-449.

Helm, D. and T Jenkinson (1998). "Introducing Competition into Regulated Industries." In Helm, D. and T Jenkinson (1998) (eds.). Competition in Regulated Industries. London: Oxford University Press, 1-22.

Hewitt, J. and W. Hanemann (1995). "A Discrete/Continuous Choice Approach to Residential Water Demand under Block Rate Pricing." Land Economics, 71(2): 173-192.

Hines L. (1969). "The Long Run Cost Function of Water Production for Selected Wisconsin Communities", Land Economics, 45 (February), 133-140.

Hittman Associates, Inc. (1970). Price, Demand, Cost and Revenue in Urban Water Utilities. Columbia, MD. NTIS PB 195 929.

Hirschleifer, J., J. de Haven and J. Milliman (1960). Water Supply: Economics Technology and Policy. Chicago: University of Chicago Press.

Hobbs, B. and G. Beim (1988). "Analytical Simulation of Water System Capacity Reliability, Modified Frequency-Duration Analysis." Water Resources Research, 24(9), 1431-1444.

Hogarty, T. F. and R. J. Mackey (1975). "The Impact of Large Temporary Rate Changes on Residential Water Use". Water Resources Research, 11(6): 791-794.

Holden, M. (1966). "Pollution Control as a Bargaining Process: An Essay on Regulatory Decision Making". Ithica: Cornell University, Water Resources Centre, Publication No. 9.

Holcombe, R. (1992). "Revolving Fund Finance." Public Budgeting and Financing, 12 (3) 50-66.

Hotelling H. (1938). "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates." Econometrica, 6(7), pp. 242-269.

Howe, C. and F. Linaweaver. (1967). The Impact of Price on Residential Water Demand and Its Relation to System Design and Price Structure. Water Resources Research, 3(1): 13-32.

Howe, C. (1982). "The Impact of Price on Residential Water Demand: Some New Insights." Water Resources Research, 18(4): 713-716.

International Joint Commission (2000). Protection of the Waters of the Great Lakes: Final Report to the Governments of Canada and the United States. Ottawa and Washington.

Jones, V. and J. Morris. (1984). "Instrumental Price Estimates and Residential Water Demand." Water Resources Research, 20(2): 197-202.

Kahn, A. (1988). The Economics of Regulation: Principles and Institutions, Cambridge, MA: MIT Press.

Katzman, M. (1977). "Income and Price Elasticities of Demand for Water in Developing Countries". Water Resources Bulletin, 13(1): 47-55.

Kellow R. (1970). A Study of Water Use in Single Dwelling Residences in the City of Calgary, AB. Edmonton: University of Alberta, Department of Economics and Rural Sociology, Unpublished Masters Thesis.

Kim, H. (1987). "Economies of Scale in Multi-Product Firms: An Empirical Analysis" Economica 54: 185-206.

Kim H. (1995). "Marginal Cost and Second-Best Pricing for Water Services", Review of Industrial Organization, 10: 323-338.

Kim Y. and R. Clark (1988). "Economies of Scale and Scope in Water Supply." Regional Science and Urban Economics. 18, 479-502.

Kneese, A. (1977). Economics and the Environment. London: Penguin Books.

Kneese, A. and B. Bower (1968). Managing Water Quality: Economic, Technology, Institutions. Baltimore: Johns Hopkins Press.

Kuhn, T. (1970). The Structure of Scientific Revolutions. Chicago: The University of Chicago.

Laffont, J. and J. Tirole (1993). A Theory of Incentives in Procurement and Regulation. Cambridge, MA: MIT Press.

Lanoie, P., M. Thomas and J. Fearnley (1998). "Firms' Responses to Effluent Regulations: Pulp and Paper in Ontario, 1985-1989." Journal of Regulatory Economics. 13:103-120.

Larson, B. O. and H. E. Hudson Jr. 1951. Residential Water Use and Family Income. Journal, American Water Works Association, 43(7): 605-611.

Lewis, A. (1946) "Fixed Costs." Economica, November, 231-258.

Linaweaver, F.P., and C. Clark (1964). "Costs of Water Transmission." Journal, American Water Works Association, 56 (12), 1549-1560.

Littlechild, S. (1983). Regulation and British Telecom's Profitability." London: Department of Industry.

Livernois J. (2002). The Economic Costs of the Walkerton Crisis. Toronto: The Walkerton Inquiry.

Löf, G. and A. Kneese (1968). The Economics of Water Utilization in the Beet Sugar Industry. Baltimore: Resources for the Future.

- Lomborg, B. (2002). The Skeptical Environmentalist. London: Cambridge University Press
- Lyon, T. (1991). "Regulation with 20:20 Hindsight: Heads I Win, Tails You Lose?" RAND Journal of Economics, 20 581-595.
- MacLaren, J.. (1985). "Municipal Water Works and Wastewater Systems". Ottawa-Hull: Inquiry on Federal Water Policy. Research Paper #3.
- MacNeill, J. (1971). Environmental Management. Ottawa: Information Canada.
- McNeill R. (1989). "Economic Theory of Water Pricing." Environment Canada, Vancouver, BC: Inland Waters Directorate, Pacific and Yukon Region, unpublished paper.
- McNeill, R. and D.M. Tate (1990). "Guidelines for Municipal Water Pricing". Ottawa-Hull: Environment Canada, Inland Waters Directorate, Social Science Research Series 25.
- Mahler, K. (1974). Environmental Economics: A Theoretical Inquiry. Baltimore: Johns Hopkins Press.
- Maidment, D. and S. Miaou, 1986. "Daily Water Use in Nine Cities." *Water Resources Research*, 22(6): 845-885
- Male, J., Willis, F. Babin, and C. Shillito, (1979). Analysis of the Water Rate Structure as a Management Option for Water Conservation. Amherst, MA: University of Massachusetts, Water Resources Research Center, Publication 112.
- Mann, P. (1981). "Water Service: Regulation and Rate Reform." Columbus, OH: National Regulatory Research Institute, Occasional Report No. 4.
- Manne, A. (1961). "Capacity Expansion and Probabilistic Growth." Econometrica 29 (4), 632-649.
- Mcguire, R. And R. Ohsfeldt (1986). "Public Versus Private Water Delivery: A Critical Analysis of a Hedonic Cost Approach", Public Finance Quarterly, 14(3): 339-350.
- Menard, A. and S. Saussier (2000). "Contractual Choice and Performance: The Case of Water Supply in France." Revue D'Economie Industrielle, 92 (2&3): 385-04,
- Merrett, S. (2000). "Industrial Effluent Policy Economic Instruments and Environmental Regulation." Water Policy 2(2): 201-211
- Mohring, H. (1970). "The Peak Load Pricing Problem with Increasing Returns and Pricing Constraints." American Economic Review, 60(4): 693-705.
- Moncur, J., (1987). Urban Water Pricing and Drought Management. Water Resources Research, 23(3): 393-398.
- Moore, F. (1959). "Economies of Scale: Some Statistical Evidence." Quarterly Journal of Economics, (May), pp. 232-245
- Morana, C., and J. Sawkins (2000). "Regulatory Uncertainty and Share Price Volatility: The English and Welsh Water Industry's Periodic Price Review." Journal of Regulatory Economics, (17) 87-100.
- Moranu-deLoë, L. (1997). "Privatizing Water Supply and Sewage Services in Ontario." Canadian Water Resources Association, Water Resource Management News and Information. Available at <http://www.cwra.org.news/arts/privatization.html>.

Morgan, W. (1973). "A Time Series Demand for Water Using Micro Data and Binary Variables." Water Resources Bulletin, 10(4): 697-702.

Morgan, W. (1982). "Water Conservation Kits: A Time Series Analysis of Conservation Policy." Water Resources Bulletin, 18(6): 1039-1042.

Morgan, W. and J. Smolen, (1976). "Climatic Indicators in the Estimation of Municipal Water Demand." Water Resources Bulletin, 12(3): 511-518.

Morris, J. and V. Jones, (1980). Water for Denver: An Analysis of the Alternatives. Environmental Defense Fund, Inc. Denver, CO.

Muir, T. and A. Sudar (1986). "Toxic Chemical in the Great Lakes Ecosystem." Burlington: Environment Canada, Discussion Paper.

Nauges, C. and A. Thomas (2000) "Privately Operated Water Utilities, Municipal Price Negotiation, and the Estimation of Residential Water Demand: The Case of France" Land Economics 76(1): 68-86.

Ng, T. and N. Weisser (1974). "Optimal Pricing with a Balanced Budget Constraint-The Case Of the Two Part Tariff." Review of Economic Studies, 41(July), 337-345.

Nguyen, D. (1976). "The Problem of Peak Loads and Inventories." Bell Journal of Economics, 7(1), PP. 232-241.

Nieswiadomy, M. (1992). "Estimating Urban Residential Water Demand: Effects of Price Structure, Conservation, and Education." Water Resources Research, 28(3): 609-15.

Nieswiadomy, M. and D. Molina (1989). "Comparing Residential Water Demand Estimates under Decreasing and Increasing Block Rates Using Household Data." Land Economics, 65(3): 280-89.

Nordin, J. (1976) "A Proposed Modification of Taylor's Demand Analysis: Comment." Bell Journal of Economics, 7(2): 719-721.

Novak, F. (1995). "Le prix de l'eau." Economica.

Ordover, J. and C. Panzar (1979). "On the Non-Existence of Pareto Superior Outlay Schedules" Bell Journal of Economics, 10, 351-354..

OECD (Organization for Economic Cooperation and Development) (1984). Economic Instruments: Review and Outlook. Paris: OECD Environment Committee, Group of Economic Experts, ENV/ECO/CI/83.9.

OECD (Organization for Economic Cooperation and Development) (1997). Evaluating Economic Instruments for Environmental Policy. Paris.

OECD (Organization for Economic Cooperation and Development) (1999). The Price of Water: Trends in OECD Countries. Paris: OECD.

OFWAT (1999). Annual Report. London.

Orlob, G., R. Lindorf (1958). "Cost of Water Treatment in California", Journal of the American Water Works Association, 50(1), 45-55.

- Ontario, (1984). Water Quantity Resources of Ontario. Toronto: Ministry Of Natural Resources.
- Ontario Office of Privatization (1998). Fact Sheet: Ontario Clean Water Agency. Toronto.
- Opaluch, J. (1982). "Urban Residential Water Demand in the United States: Further Discussion." Land Economics, 58: 225-227.
- Pearse, P.H., F. Bertrand, and J.W. MacLaren (1985). Currents of change: final report of the Inquiry on Federal Water Policy. Ottawa-Hull, Environment Canada, Catalogue No. En 37-71/1985-1E.
- Pint, E. (1999). "Household Responses to Increased Water Rates During the Californian Drought" Land Economics, 75(2) 246-66.
- Pollak R. and Wales T. (1992). "Demand System Specification and Estimation". Oxford: Oxford University Press.
- Polzin, P. (1984). "The Specification of Price in Studies of Consumer Demand under Block Price Scheduling: Additional Empirical Evidence." Land Economics, 60(3): 305-309.
- Pope, R. and J. Stepl, and J. Lytle, (1975). Effects of Price Change Upon the Domestic Use of Water Over Time. Water Resources Research Institute. Clemson University, South Carolina.
- Primeaux, W. and K. Hollman, (1973). "Price and Other Selected Economic and Socio-Economic Factors as Determinants of Household Water Consumption." In Water for the Human Environment. Proceedings of the First World Congress on Water Resources, vol. 3 Champaign, IL: International Water Resources Association.
- Pressman, I. (1970). "A Mathematical Formulation of the Peak Load Pricing." Bell Journal of Economics, 1(2), 304-22.
- Raffiee, K., R. Narayanan, T. Harris, D. Lambert and J. Collins (1993). "Cost Analysis of Water Utilities: A Goodness-of-Fit Approach", Atlantic Economic Journal, 21(3): 18-29.
- Rankin, M. T. (1991). "Despoiling a River: Can the Law Help to Sustain the Fraser River?". in Dorcey, A. Sustainable Development in the Fraser Basin. Vancouver: Westwater Research Centre.
- Rees, R. (1984). Public Enterprise Economics. New York: St. Martins Press.
- Renwick, M. and S. Archibald (1998). "Demand Side Management Policies for Residential Water Use: Who Bears the Conservation Burden?" Land Economics, 74(3): 343-59.
- Renzetti, S. (1992a). "Estimating the Structure of Industrial Water Demands: The Case of Canadian Manufacturing." Land Economics, 68(4): 396-404.
- Renzetti, S. (1992b). "Evaluating the Welfare Effects of Reforming Municipal Water Prices", Journal of Environmental Economics and Management, 22(2), 147-163.
- Renzetti, S. (1999). "Municipal Water Supply and Sewage Treatment: Costs, Prices and Distortions" Canadian Journal of Economics, 32(2): 688-704.
- Renzetti, S. (2000). "Environmental Conditions and the Cost of Municipal Water Supply." St. Catherines, ON: Brock University, unpublished manuscript.

Russell, C. (1973). Residuals Management in Industry: A Case Study of Petroleum Refining. Baltimore: Resources for the Future.

Russell, C. and B. Shin (1996). "Public Utility Pricing: Theory and Practical Limitations." In Darwin Hall (ed.) Marginal Cost Rate Design and Wholesale Water Markets: Advances in the Economics of Environmental Resources, vol. 1. Greenwich, Conn: JAI Press.

Saal, D. and D. Parker (2000). "The Impact of Privatization and Regulation on the Water and Sewerage Industry in England and Wales: A Translog Function Model." Managerial and Decision Economics, 21: 253-268.

Saleth, R.M. and A. Dinar (1997). "Satisfying Urban Thirst: Water Supply Augmentation and Pricing Policy in Hyderabad City, India." Washington D.C.: The World Bank. World Bank Technical Paper 395.

Sawkins, J. (1996). "Balancing Multiple Interests in Regulation: An Event Study of the English and Wales Water Industry." Journal of Regulatory Economics, 9: 249-268.

Scarato, R., (1969). "Time Capacity Expansion of Urban Water Systems." Water Resources Research, 5(October), 929-936.

Scherer, F. and D. Ross (1990). Industrial Market Structure and Economic Performance. Boston, MA: Houghton Mifflin and Company,

Schultze, C.W.(1977). "The Public Use of Private Interest." The Godkin Lectures. Cambridge, MA: Harvard University Press.

Schumpeter, J. (1942). Capitalism, Socialism, and Democracy. New York: Harper and Row.

Seidel, H. and E.. Baumann. 1957. A Statistical Analysis of Water Works Data for 1955. Journal, American Water Works Association, 42(12): 1531-1566.

Sewell, W. R. D. and L. Roueche, (1974). "Peak Load Pricing and Urban Water Management: Victoria, B.C., A Case Study." Natural Resources Journal, 14(3): 383-400.

Sheperd, W. (1998). "Problems of Creating Effective Competition." In Gabel, D. and D. Weiman (eds.), Opening Networks to Competition: The Regulation and Pricing of Access. Boston: Kluwer Academic Publishers, 49-71.

Shin, J. (1985). Perception of Price when Information is Costly: Evidence from Residential Electricity Demand. Review of Economics and Statistics, 67(4): 591-598.

Sierra Legal Defense Fund (2001). Waterproofing Canada. Vancouver.

Sims, W. (1979). "The Response of Firms to Pollution Charges." Canadian Journal of Economics 12(1): 57-74.

Solley, W., R. Pierce and H. Perlman (1998). "Estimated Use of Water in the United States in 1995." Washington, D.C.: United States Government Printing Office, U.S. Geological Survey Circular 1200.

Solow, R. (1957). "Technical Change and the Aggregate Production Function". Review of Economics and Statistics, 39: 312-320.

Steiner, P. (1957). "Peak Loads and Efficient Pricing." Quarterly Journal of Economics. 71(4), 585-610

Stephan, G. (1988). "Economic Impact of Emission Standards: A Computational Approach to Wastewater Treatment in Western Europe." in D. Bös, M. Rose and C. Seidl, (eds.) Welfare and Efficiency in Public Economics. Berlin: Springer-Verlag pp. 401-22.

Tate, D. (1984). "Canadian Water Management: A One-Armed Giant." Canadian Water Resources Journal, 9, #3, pp. 1-7.

Tate, D. (1990). "Water Demand Management in Canada: A State of the Art Review." Ottawa-Hull: Environment Canada, Inland Waters Directorate. Social Science Series No. 23 .

Tate, D. (2001). A Brief Overview of Water Quantity and Related Issues In Ontario. Toronto: The Walkerton Inquiry.

Tate, D. and J. Harris (1999). Water Demands In The Canadian Section Of The Great Lakes Basin, 1972 – 2021. Ottawa: Canadian Section, International Joint Commission.

Tate, D. and D. Lacelle (1995). "Municipal Water Rates in Canada: Current Practices and Prices, 1991." Ottawa-Hull: Environment Canada, Inland Waters Directorate, Social Science Series, 30.

Tate, D. and D. Scharf, (1995). "Water Use in Canadian Industry: 1991." Ottawa-Hull: Department of the Environment, Canadian Wildlife Service. Social Science Series, 31.

Taylor, L. (1975) "The Demand for Electricity: A Survey." Bell Journal of Economics. 6(1): 74-110.

Tchobanoglous, G. and E. Schroeder (1987). Water Quality. New York: Addison Wesley.

Teeple, R. and D. Glycer (1987b). "Cost of Water Delivery Systems: Specification and Ownership Effects", Review of Economics and Statistics, 69(3): 399-408.

Terkla, D. (1984). "The Efficiency Value of Effluent Tax Revenues." Journal of Environmental Economics and Management, 11: 107-123.

Tietenberg, T. (1985). Emissions Trading: An Exercise in Reforming Pollution Policy. Washington, D.C.: Resources for the Future.

Tietenberg, T. (1996). Environmental and Natural Resource Economics. Boston: Scott, Foresman and Co.

Train, K. (1994). Optimal Regulation: The Economic Theory of Natural Monopoly., Cambridge, MA: MIT Press.

Turnovsky, S. J. (1969). "The Demand for Water: Some Empirical Evidence of Consumers' Response to a Commodity Uncertain in Supply." Water Resources Research, 5(2): 350-361.

Turvey, R. (1969). "Marginal Cost." Economic Journal, 79 (6), 282-299.

Turvey, R. (1976). "Analyzing the Marginal Cost of Water Supply." Land Economics, 52(5), 158-168.

U.S. EPA (1995). The Clean Water State Revolving Fund: Financing America's Infrastructure – A Report of Progress. Washington: EPA 832 R-95-001.

Varian, H. (1992). Microeconomic Analysis. New York: W.W Norton.

- Vickery, W. (1948). "Some Objections to Marginal Cost Pricing", The Journal of Political Economy, 56(June), pp. 218-238.
- Vickery, W. (1955). "Some Implications of Marginal Cost Pricing for Public Utilities." American Economic Review, 45(5), 605-620.
- Vickery, W., 1971. "Responsive Pricing of Public Utility Services", Bell Journal Of Economics, 2(1), 337-346.
- Ware, J. and R. North (1967). "The Price and Consumption of Water for Residential Use in Georgia." , Atlanta, GA: Georgia State University, Bureau of Business and Economic Research, School of Business Administration,.
- Water Supply and Sanitation Collaborative Committee (1998). Water Demand Management and Conservation. (by Forde, L., D. Tate, and S. Arlosoroff). Geneva.
- Weber, J. (1989). "Forecasting Demand and Measuring Price Elasticity." Journal of the American Water Works Association, 81(5): 57-65.
- Whittington, D., D. Lauria, K. Choe, J. Hughes, V. Swarna and A. Wright (1993). "Household Sanitation in Kumasi, Ghana: A Description of Current Practices, Attitudes and Perceptions." World Development, 21(5): 733-748.
- Williamson, O. (1966). "Peak Load Pricing and Optimal Capacity under Indivisibility Constraints." American Economic Review, 56 (9), pp. 810-27.
- Williamson, O. (1976). "Franchise Bidding for Natural Monopolies – In General and with respect to CATV." Bell Journal of Economics, 7: (1), 73-104.
- Williamson, O. (1985). The Economic Institutions of Capitalism." New York: The Free Press.
- Willig, R. (1978). "Pareto-Superior Nonlinear Outlay Schedule." Bell Journal of Economics. 9(2), pp. 56-69.
- Wiseman, J., 1959, "The Theory of Public Utility Price - An Empty Box", Oxford Economic Papers, 9(2), 56-74.
- Wong, S. T. (1972). "A Model of Municipal Water Demand: A Case Study of Northeastern Illinois." Land Economics, 48(1): 34-44.
- World Bank (1994). World Development Report: Infrastructure for Development. Toronto: Oxford University Press.
- Young, R. (1971). "Price Elasticity of Demand for Municipal Water: A Case Study of Tucson, Arizona." Water Resources Research, 9(4): 1068-1072 9(4): 1068-1072.