CSIS Discussion Paper No. 64

Rate-of-return and Price-cap Regulations for Congested Urban Railways

Yukihiro Kidokoro*

Abstract: Using a simple model of commuter railways where congestion exists, we show that the regulatory shift from rate-of-return regulation to price-cap regulation does not cure the congestion problem in Japanese urban railways. We next consider methods to correct it, and show the following results: (i) PC regulation, in which the cap is made contingent on congestion, can correct the congestion without distorting cost-reducing efforts, (ii) PC regulation, in which the cap depends on investment in transportation capacity, can also correct the congestion but distorts cost-reducing efforts.

Keywords: price-cap regulation; rate-of-return regulation; congestion; railways

---

* Center for Spatial Information Science, University of Tokyo

Mailing Address: Faculty of Economics, University of Tokyo, 7-3-1, Hongo, Bunkyo, Tokyo, 113-0033, Japan

Tel & Fax: +81-3-5841-5641, e-mail: kidokoro@csis.u-tokyo.ac.jp

I would like to thank Yoshitsugu Kanemoto for helpful suggestions and discussions. I owe special thanks to Hajime Miyazaki for his extensive comments. I would also like to thank Ken Brevoort and Tatsuo Hatta for their remarks. Of course, I am solely responsible for any errors and omissions.

1 Introduction

Urban railways, including public and private railways, play important roles as part of Japan's urban transit systems. For example, railways in the Tokyo metropolitan area provides 55% of travel needs in 2001. Urban railways in Japan, however, have a serious problem; extreme congestion during rush hours. The congestion rate, which is the ratio of the number of users to a railway's nominal capacity, is about 200% in almost every line in the Tokyo metropolitan area. Typically, Tokyo railway firms have been operating their lines with very short headway and very long train configurations during peak time. Therefore, peak-time congestion cannot be relieved without constructing new lines or enhancing the current double-track lines to four-track lines.

The purpose of this paper is to investigate whether the congestion problem stated above is resolved or not by the regulatory reform from rate-of-return (ROR) regulation[1] to price-cap (PC) regulation[2], which is now under way in many industrialized countries, and to propose the regulatory method that is consistent with the relief of the congestion. To deal with the problem, we build a simple spatial land-use model of commuter railways and analyse the effects that are caused by the regulatory shift from ROR regulation to PC regulation. The reason why we employ a spatial land-use model is that the bottleneck input in enhancing transportation capacity from the current double-track to four-track lines in Japan is railroad right-of-way.

This paper shows that the regulatory shift to PC regulation does not cure the congestion problem although it corrects the distortion in input mix under ROR regulation. We next focus on modifications of PC regulation and obtain the following results: (1) PC regulation with a cap contingent on transportation quality, which is the inverse of the congestion rate, can relieve the congestion without distorting cost-reducing efforts. (2) PC regulation, in which the cap is made contingent on investment, can also correct the congestion but distorts cost-reducing efforts.

---

[1] See, for example, Averch and Johnson (1962) and Train (1991) for the theory of ROR regulation.

[2] See, for example, Littlechild (1983) and Armstrong et al. (1994) for the theory of PC regulation.

The structure of the paper is as follows. In Section 2, we set up a simple spatial model. In Section 3, we obtain the results of the first best case as a benchmark. Section 4 analyzes ROR regulation and Section 5 does PC regulation. In Section 6, modifications of PC regulation are considered. Section 7 concludes our analysis.

2 Model

Our model is a variant of the model outlined in Kidokoro (1998), that is based on an open-city and absentee-landlord model from urban economics literature[3]. Consider a residential city of fixed size $\overline{H}$, which is connected to the central business district (CBD) by a railway. The utility level outside the city is given by $\overline{u}$. Residents and potential residents can freely move out of and into the residential city. As a result, residents attain exactly $\overline{u}$ in equilibrium. For simplicity, we assume that there are no transportation costs within the residential city. Consequently, transportation fares are uniform within the residential city. This assumption implies that we disregard locational differences within the residential city.

All residents in this city, N in number, are assumed to use a railway line to make a round trip to and from the CBD every day. Moreover, they are assumed to commute to the CBD in peak time when they are on duty and to go to the CBD to dine or shop in off-peak time when they are off duty. If all the residents had the same pattern of on-duty days and off-duty days, they would use a railway line in peak time on their on-duty days and, consequently, off-peak rail demand would not exist on their on-duty days. In order to create both peak-time and off-peak-time demand daily, we further assume that the pattern of on-duty days and off-duty days differs resident by resident. For example, resident A, who works in an electronics company, may work from Monday to Friday, while resident B, who works in a supermarket, may work from Monday to Wednesday, Saturday and Sunday. The upshot is that the daily peak-time users of a railway are commuters who are on duty and the daily off-peak time

[3] See, for example, Kanemoto (1980).

users are non-commuters who are off duty.

A railway firm supplies transportation capacity Q daily. Q is nominal in the sense that the actual number of daily users of a railway, which is the sum of peak-time users and off-peak-time users, is not Q but N. We here define the daily average congestion rate as $\frac{N}{Q}$. To allow analytical simplicity, we use the inverse of the daily average congestion rate, $q \equiv \frac{Q}{N}$, as the daily average quality of transportation service. For example, if a railway firm transports many residents with low transportation capacity every day, the daily average congestion rate is high and hence the daily average quality of transportation service is low. To differentiate the peak time from the off-peak time, the peak-time quality of the transportation service, q^{peak}, is defined as $q^{peak} \equiv \frac{Q^{peak}}{N^{peak}}$, and the off-peak-time congestion rate, $q^{offpeak}$, is defined as $q^{offpeak} \equiv \frac{Q - Q^{peak}}{N - N^{peak}}$, where Q^{peak} and N^{peak} are daily peak-time transportation capacity and daily peak-time users of a railway, respectively.[4] For the sake of simplicity, we assume that transportation capacity and the users of a railway line in peak time are functions of their daily values, i.e., $Q^{peak} = f(Q)$ and $N^{peak} = g(N)$ and that $q^{peak} = \frac{Q^{peak}}{N^{peak}} = \frac{f(Q)}{g(N)} = \zeta\left(\frac{Q}{N}\right) = \zeta(q)$. Thus, q^{peak} is a function of q and can be written as $q^{peak}(q)$.

We assume that all residents are homogeneous with a quasi-linear and additively separable utility function, $U(z, h, q^{offpeak}, q^{peak}) = z + u(h) + v(q^{offpeak}, q^{peak})$, where z is the composite consumer good, including housing, whose price is normalized at one, and h is residential lot size. Since the

[4] We implicitly assume that daily peak-time transportation capacity and the number of daily peak-time users of a railway are the same for all days and disregard the daily variance in them. As a result, the quality of transportation service for peak and off-peak times also becomes day-invariant.

off-peak congestion rate is sufficiently low even in Japanese urban railways, we disregard it and simplify the utility function to $U(z,h,q^{peak}) = z + u(h) + v(q^{peak})$ to focus on peak-time congestion. That is, in our analysis, the off-peak quality of transportation service is assumed not to affect the utility level of residents. This simplification of the utility function enables us to conduct a simulation based on the actual estimate of $v(q^{peak})$. We assume that $u(h)$ and $v(q^{peak})$ are strictly increasing and strictly concave: $u'(h) > 0$, $u''(h) < 0$, $v'(q^{peak}) > 0$, $v''(q^{peak}) < 0$. As is well known, the income effects are zero under the quasi-linearity assumption[5] and cross elasticities are zero under the separability assumption. As a result, this form of the utility function yields a demand function for land that depends only on land rent, which simplifies our analysis.

Each resident solves his or her utility maximization problem subject to a budget constraint, $z + Rh + t = \overline{w}$, where R, t, and $\overline{w}$ denote land rents, transportation fares, and the fixed income of a resident, respectively. Maximizing the utility function, $U(z,h,q^{peak}) = z + u(h) + v(q^{peak})$, under the budget constraint, $z + Rh + t = \overline{w}$, yields $u'(h) = R$. Inverting this function yields the demand function for land:

$$h = h(R) \equiv u'^{-1}(R).$$

Following the usual procedure in the urban economics literature, we derive the bid rent function, which gives the maximum possible rent, providing utility level $\overline{u}$. The bid rent function is:

$$R(\overline{y} - t, q^{peak}) = \max_{\{z,h\}}\{\frac{\overline{w} - t - z}{h} : z + u(h) + v(q^{peak}) \geq \overline{u}\},$$

where $\overline{y}(\equiv \overline{w} - \overline{u})$ is real income. This function satisfies $R_I = \frac{1}{h} > 0$, $R_t = -\frac{1}{h} < 0$, and $R_{q^{peak}} = \frac{v'(q^{peak})}{h} > 0$, where $I \equiv \overline{y} - t$. Hereafter, the subscripts denote partial derivatives, unless otherwise noted.

---

[5] See, for example, Varian (1992).

Substituting the bid rent function into the demand function for land yields a lot size function from which we eliminate the land rent, R:

$$\hat{h}(I, q^{peak}) = h(R(I, q^{peak})).$$

This function satisfies $\hat{h}_I = \frac{h_R}{h} < 0$, $\hat{h}_t = -\frac{h_R}{h} > 0$, and $\hat{h}_{q^{peak}} = \frac{h_R v'(q^{peak})}{h} < 0$.

We assume that the railway firm uses L as railroad right-of-way within the city area $\overline{H}$. (We ignore railroad right-of-way used outside the city area.) The residential area left for housing is thus $\overline{H} - L$. We assume that railroad right-of-way can be converted without cost into residential land and vice versa. Given $\overline{u}$, the equilibrium number of residents is then defined as:

$$N(t, q^{peak}, L) = \frac{\overline{H} - L}{\hat{h}(\overline{y} - t, q^{peak})}.$$

To the railway firm, $N(t, q^{peak}, L)$ is synonymous with the equilibrium transport demand function given $\overline{u}$. The transport demand function satisfies $N_t = \frac{N h_R}{h^2} < 0$, $N_{q^{peak}} = -\frac{N h_R v'(q^{peak})}{h^2} > 0$, and $N_L = -\frac{1}{h} < 0$.

The production function, by which a railway firm supplies daily railway capacity Q, is $Q = qN = F(L, Z, e)$, where L is railroad right-of-way, Z is non-capital input, and e is the firm's efforts. We assume that $F_i \equiv \frac{\partial F}{\partial i} > 0$ and $F_{ii} \equiv \frac{\partial^2 F}{\partial i^2} < 0$, where $i = L$, Z, and e.[6] We ignore capital inputs other than railroad right-of-way for simplicity, because our focus is on railroad right-of-way, which is the bottleneck input that hinders an enhancement of railway capacity. For simplicity, the non-capital input, Z, is assumed to be the same good as the composite consumer good. The price of

[6] We disregard the problem of indivisibilities of investments by assuming that the investment continuously changes transportation capacity. This assumption at least approximately holds in reality, because it is common to enhance only a part of track capacity, e.g., terminal stations, the most congested part of a railway line, or bottleneck intersections, in order to reduce total investment costs.

the non-capital input is therefore one. We call the firm's efforts, e, cost-reducing efforts, because under the conditions of our model, the firm's efforts to expand railway capacity are included in the production function, which implies that the firm's efforts do reduce its costs, given railway capacity. A railway firm is assumed to incur monetary disutility $\varphi(e)$ from cost-reducing efforts, e, where we assume $\varphi'(e) > 0$ and $\varphi''(e) > 0$. Both the cost-reducing efforts and the production function itself are unobservable to the regulator. This means that the regulator cannot calculate the value of e based on Q, L, and Z. Thus, the regulator cannot implement the regulation based on the level of e.

A railway firm is assumed to have obtained $d\bar{H}$ (d is any number between 0 and 1) of the residential area with the price of $\bar{V}$ before a railway line was built. This assumption corresponds to the fact that railway firms in Japan own much land, the book price of which is very low, along their railway lines. The railway firm rents $d_1\bar{H}$ [7] out of $d\bar{H}$ as residential land. Thus, the railway firm has two sources of revenue: railway fares and rental income from residential land. For simplicity, the provision of residential land is the only side business in our analysis. The railway firm uses the other part of $d\bar{H}$, $d_2\bar{H} = (1 - d_1)\bar{H}$, as its railroad right-of-way. (Since the total railroad right-of-way is L, $d_2\bar{H} < L$).[8] The regulatory authority is assumed to impose regulations only on the firm's railway sector, as is the case in Japan. We also assume that many competitive absentee-landlords own the

---

[7] d_1 and d_2 are exceptions to our notation rule that the subscripts denote partial derivatives.

[8] In the ROR regulation in Japan, railroad right-of-way in the rate base is undervalued, compared with its market value. In this case, a railway firm has an incentive to use the land it owns, not as railroad right-of-way but to rent it out as residential land, i.e., to make $d_2 = 0$ if possible. This is because lower evaluation in the rate base leads to lower allowed profits of railway business, which are less than the market land rents it could obtain. However, in reality, a lot of the land owned by railway firms is now being used as railroad right-of-ways, probably due to technological factors affecting the construction of railway lines or to the nature of the tax system, issues that we do not address in this paper. Thus, in this paper, d_2 is assumed to be a fixed positive value. In considering the PC regulation and its variants, which is the main focus of this paper, this assumption is innocuous, because they have no relationship with the rate base, and thus the magnitude of d_2 does not affect the railway firms' profits.

railway firm and the residual residential area, $(1-d)\bar{H}$.

Now we can express the profits of the railway firm as:

$$\pi(t,q,L,Z,e) \equiv tN(t,q^{peak}(q),L) + R(\bar{y}\text{-}t,q^{peak}(q))d\bar{H} - Z - R(\bar{y}\text{-}t,q^{peak}(q))L - \varphi(e) .$$

In our closed-form model, social welfare is the sum of the railway firm's profits and the residential land rent that does not accrue to the railway firm. Social welfare is then:

$$SW(t,q,L,Z,e) \equiv \pi(t,q^{peak}(q),L,Z,e) + R(\bar{y}\text{-}t,q^{peak}(q))(1\text{-}d)\bar{H} .$$

We consider an infinite period, beginning in period 0 when the railway line is built. Since we set up no changes after period 0, the state in period 0 is repeated endlessly. Therefore, the railway firm keeps its choice variables, i.e., t, q, L, Z, and e, constant from period 0 onwards. The present discounted value of the railway firm's profits and social welfare from period 0 onwards, respectively, can then be written as

$$PV(\pi) \equiv \frac{\pi(t,q,L,Z,e)}{r} \tag{1}$$

and

$$PV(SW) \equiv \frac{SW(t,q,L,Z,e)}{r}, \tag{2}$$

where r is the fixed cost of capital.

3 First Best

First, as a benchmark, we obtain the first best input choices by maximising the present discounted value of social welfare, (2), subject to the production constraint:

$$F(L,Z,e) \geq qN(t,q^{peak}(q),L) . \tag{3}$$

The results are stated as Proposition 1.

Proposition 1

In the first best optimum,

(i) The marginal benefit of the investment in the non-capital input, Z, equals its marginal cost: $v'(q^{peak})q^{peak\prime}(q)F_Z = 1$.

(ii) The marginal rate of technical substitution between the non-capital input, Z, and railroad right-of-way, L, equals their relative price: $\frac{F_L - qN_L}{F_Z} = R - tN_L$.

(iii) The marginal rate of technical substitution between the non-capital input, Z, and cost-reducing efforts, e, equals their relative price: $\frac{F_e}{F_Z} = \varphi'(e)$.

The proof is in Appendix 1. In result (i), $v'(q^{peak})q^{peak\prime}(q)F_Z$ shows the marginal benefit of the investment in the non-capital input. An additional investment in the non-capital input expands transport capacity by F_Z, which increases peak-time transportation quality by $q^{peak\prime}(q)F_Z$. The marginal benefit from peak-time transportation quality is $v'(q^{peak})$, and consequently, $v'(q^{peak})q^{peak\prime}(q)F_Z$ shows the marginal benefit of an increase in peak-time transportation quality caused by the investment in the non-capital input, i.e., the marginal benefit of the investment in the non-capital input. Result (i) shows that there is no distortion in the choice of non-capital input, because the marginal benefit of the investment in the non-capital input equals its marginal cost.

In result (ii), $\frac{F_L - qN_L}{F_Z}$ shows the marginal rate of technical substitution between the non-capital input, Z, and railroad right-of-way, L, which has a different form from the standard one. The reason is that investment in railroad right-of-way increases railway capacity in two ways. First, the investments directly increase railway capacity by F_L. Second, they decrease the land available for housing, and consequently decrease the number of residents (= users). (Recall that city size, $\overline{H}$, is fixed.) This decrease in the number of users, N_L, virtually expands railway capacity by $-qN_L$, because a user uses the capacity by q on average. The effect of investments in railroad right-of-way on railway capacity is thus $F_L - qN_L$. In the end, the marginal rate of technical substitution between Z and L is $\frac{F_L - qN_L}{F_Z}$. The decrease in the number of users also reduces a railway firm's fare revenue by $-tN_L$, which makes the opportunity cost of railroad right-of-way $R - tN_L$. The relative price of it to non-capital input is then $R - tN_L$. (Recall that the price of Z is normalised at one.) Result (ii) states that the choice of railroad right-of-way, compared to the choice of non-capital input, is not distorted, because the marginal rate of technical substitution equals the relative price.

In Result (iii), $\varphi'(e)$ is the marginal disutility of cost-reducing efforts, e. $\varphi'(e)$ can then be interpreted as the marginal cost of the railway firm's cost-reducing efforts, and the relative price of e to Z. Results (iii) states that the choice of cost-reducing efforts is not distorted, compared to the choice of non-capital input.

These three conditions guarantee that input choices are optimal. Thus, by comparing these conditions with those under the regulatory methods we focus on hereafter, we can know the influences of the regulatory methods on railway firms' input choices.

4 ROR Regulation

To compare the case of ROR regulation with that of PC regulation later, we here consider a railway firm's behaviour under ROR regulation, which can be formulated as

$$\rho \times (RateBase) \geq tN(t, q^{peak}(q), L) - Z ,$$

where ρ is the allowed rate of return. Rate base can be written as $V(L - d_2\bar{H}) + \bar{V}d_2\bar{H}$, where V denotes land price, when the regulation is based on book value, while VL when the regulation is based on market value. (Recall that the purchase price of $d_2\bar{H}$ is $\bar{V}$.) Combining the case of book value with that of market value, we formulate rate base as $V(L - \beta d_2\bar{H}) + \beta\bar{V}d_2\bar{H}$, where $\beta = 1$ shows book-value-based ROR regulation and $\beta = 0$ shows market-value-based ROR regulation.

In the end, ROR regulation is

$$\rho\left\{V(L-\beta d_2\bar{H})+\beta\bar{V}d_2\bar{H}\right\}\geq tN(t,q^{peak}(q),L)-Z\,. \tag{4}$$

The maximisation of the present discounted value of the railway firm's profits, (1), subject to the production constraint, (3), and ROR regulation, (4), yields basically the same results as Kanemoto and Kiyono (1993,1995). We summarise the results as Proposition 2.

Proposition 2

Under binding ROR regulation,

(i) The marginal benefit of the investment in the non-capital input, Z, equals its marginal cost: $v'(q^{peak})q^{peak\prime}(q)F_Z=1$.

(ii) If the allowed rate of return, ρ, is lower than the true cost of capital, r, then the marginal rate of technical substitution between the non-capital input, Z, and railroad right-of-way, L, exceeds their relative price. If $\rho>r$, then the usual AJ result holds. If $\rho=r$, then the marginal rate of technical substitution equals their relative price: $\dfrac{F_L-qN_L}{F_Z}\gtreqless R-tN_L$ as $\rho\lesseqgtr r$.

(iii) The marginal rate of technical substitution between the non-capital input, Z, and cost-reducing efforts, e, exceeds their relative price: $\dfrac{F_e}{F_Z}>\varphi'(e)$.

The proof is stated in Appendix 1. Result (i) is the same as Proposition 1-(i). That is, ROR regulation does not distort a railway firm's choice of non-capital input[9]. Under ROR regulation, the railway firm is allowed to raise its price when investing in non-capital input. Therefore, the railway firm has no incentive to lower the investments in non-capital input.

---

[9] If the utility function is not quasi-linear or additively separable, a railway firm's choice of non-capital input is distorted compared with the first best case: non-capital input is more or less than the optimal level. For this distortion, see Spence (1975).

Result (ii) states that the usual AJ effect[10] should be modified when railway firms run profitable side businesses. Since there are profits from side businesses, the regulator can set the rate of return, ρ, below the true cost of capital, r, in which case, the railway firm's return from the investment in railroad right-of-way is lower than its costs. This low rate of return leads to underinvestment in railroad right-of-way, and reverses the AJ result: the marginal rate of substitution between railroad right-of-way, L, and non-capital, Z, exceeds their relative price. Kanemoto and Kiyono (1993, 1995) point out that this underinvestment due to the reversed AJ effect is one of the main reasons for the congestion of urban railways in Japan's large cities. If $\rho > r$, the normal AJ result holds. If $\rho = r$, the marginal rate of substitution equals the relative price, and consequently the input choice between L and Z is not distorted.

Result (iii) shows that ROR regulation distorts cost-reducing efforts. If a railway firm reduce its costs by its effort, its profits increase. However, given this increase in profits, a firm's actual rate of return goes up and becomes higher than the allowed rate of return. In this case, the firm must reduce its revenue to meet ROR regulation. This implies that the railway firm cannot capture the increase in profits that cost reduction yields, and has weaker incentives for cost-reduction. Thus, under ROR regulation, the railway firm operates inefficiently.

Actually, constructing a new railway line entails huge development profits. That is, land prices go up substantially after the construction of new railway lines. Railroad right-of-way is revalued after construction of railway lines under market-value-based ROR regulation, while it is never revalued under book-value-based ROR regulation. The difference between market value and book value of railroad right-of-way is very large in Japan even after the burst of the Bubble economy in the early 1990s. For major railway companies operating in the Tokyo metropolitan area, the book value of railroad right-of-way is less than 10 % of its market value in 1993 prices. Is the lack of revaluation of railroad right-of-way under book-value-based ROR regulation hurt cost-reducing incentives? The

[10] See Averch and Johnson (1962).

following corollary answers the question.

Corollary 1

When the allowed rate of return, ρ*, equals the true cost of capital,* r*,*

(i) under book-value-based ROR regulation, incentives for cost reduction exist as long as a railway firm engages in side businesses, i.e., $d_1 > 0$.

(ii) Under market-value-based ROR regulation, incentives for cost reduction exist as long as the railway firm owns land, i.e., $d > 0$.

The proof is stated in Appendix 1. Note that cost-reducing incentives are stronger under market-value-based ROR regulation than under book-value-based ROR regulation. If the regulation is based on the book value of land, the cost-reducing incentives for the railway firm with no side business, i.e., $d_1 = 0$, disappear when the allowed rate of return, ρ, equals the true cost of capital, r. A railway firm with side businesses, i.e., $d_1 > 0$, retains these incentives even when $\rho = r$ because the lower fare brought about by cost reductions leads to higher residential land rent and hence larger profits from side businesses.

If the regulation is based on the market value of land, however, even a railway firm that owns only railroad right-of-way, (i.e., $d_1 = 0$ and $d_2 > 0$) keeps the incentives for cost reduction in the case of $\rho = r$. In this case, the lower fare from reduced costs yields higher land rent, which leads to higher land prices. With revaluation under market-value-based ROR regulation, the higher land price makes the rate base larger. The larger rate base then yields larger profits from the railway business. Thus, the revaluation of railroad right-of-way after construction of railway lines in the case of market-value-based ROR regulation strengthen the cost-reducing incentives.

The above analyses show that (1) ROR regulation may increase the level of congestion due to underinvestment in railroad right-of-way as a result of the reversed AJ effect, (2) ROR regulation causes the distortions in incentives in cost-reducing efforts, although they are somewhat alleviated by adopting the market-value-based ROR regulation.

5 PC Regulation

What happens when the regulation shifts to PC regulation? We now focus on PC regulation, which can be written as

$$tcap \geq t\,, \tag{5}$$

where $tcap$ is the fixed ceiling price. The maximisation of the present discounted value of the railway firm's profits, (1), subject to the production constraint, (3), and PC regulation, (5) yields the railway firms' input choices under PC regulation, which are summarised as proposition 3.

Proposition 3

Under binding PC regulation,

(i) The marginal benefit of the investment in the non-capital input, Z, exceeds its marginal cost: $v'(q^{peak})q^{peak\prime}(q)F_Z > 1$.

(ii) The marginal rate of technical substitution between the non-capital input, Z, and railroad right-of-way, L, equals their relative price: $\dfrac{F_L - qN_L}{F_Z} = R - tN_L$.

(iii) The marginal rate of technical substitution between the non-capital input, Z, and cost-reducing efforts, e, equals their relative price: $\dfrac{F_e}{F_Z} = \varphi'(e)$.

The proof is in Appendix 1. Result (i) shows that PC regulation leads to underinvestment in non-capital input. This distortion occurs for the following reason. As long as the transportation fare is suppressed by PC regulation, the railway firm cannot raise the fare even if it invests in non-capital input. Investment in non-capital input is therefore discouraged. Results (ii) and (iii) state the merits of PC regulation; under PC regulation, the railway firm employs the optimal input mix between non-capital input, Z, railroad right-of-way, L, and cost-reducing efforts, e, because it regulates the ceiling price only. That is, given the railway firm's investment level of non-capital input, choices of railroad right-of-way and cost-reducing efforts are not distorted. This result implies that the tendency for an inefficient operation under ROR regulation disappears under PC regulation.

The above analysis of PC regulation demonstrates that compared to ROR regulation, PC regulation contains another source of distortion in investments, which leads to congestion. Under PC regulation, non-capital input are underinvested, which also leads to underinvestment in railroad right-of-way and cost-reducing efforts, provided that the marginal rate of technical substitutions among all inputs are optimal. That is, as long as the ceiling price is binding, PC regulation causes underinvestment in *all* inputs and thus causes congestion. The congestion problem does not disappear under the regulatory shift from ROR to PC regulation.

6 Modified PC Regulation

The analyses in the last section make it clear that the regulatory shift to PC regulation does not resolve the congestion problem. We consider here modified versions of PC regulation that is consistent with the relief of the congestion.

First, let us focus on PC regulation with a cap contingent on transportation quality, which is the inverse of the congestion rate. PC regulation thus modified can be written as

$$tcap + \overline{t}(q^{peak}) \geq t, \tag{6}$$

where $\overline{t}(q^{peak})$ is the variable part of the ceiling price that depends on peak-time transportation quality, q^{peak}. We call this quality-contingent PC regulation. Maximising the present discounted value of the railway firm's profits, (1), subject to the production constraint, (3), and quality-contingent PC regulation, (6), we obtain the railway firms' input choices under quality-contingent PC regulation, which are summarised as proposition 4.

Proposition 4

Under quality-contingent PC regulation, if $\overline{t}'(q^{peak}) = v'(q^{peak})$,

(i) The marginal benefit of the investment in the non-capital input, Z*, equals its marginal cost:* $v'(q^{peak})q^{peak\prime}(q)F_Z = 1$.

(ii) The marginal rate of technical substitution between the non-capital input, Z*, and railroad right-of-way,* L*, equals their relative price:* $\dfrac{F_L - qN_L}{F_Z} = R - tN_L$.

(iii) The marginal rate of technical substitution between the non-capital input, Z*, and cost-reducing efforts,* e*, equals their relative price:* $\dfrac{F_e}{F_Z} = \varphi'(e)$.

The proof is in Appendix 1. Quality-contingent PC regulation eliminates the drawbacks, while maintaining the merits of PC regulation. Under quality-contingent PC regulation, a railway firm is allowed to set a higher price when alleviating congestion. Since the railway firm can obtain profits from its investments in railway capacity, it has no actual incentive to decrease the investments in railroad right-of-way and non-capital input and to thereby increase congestion. If $\overline{t}'(q^{peak}) = v'(q^{peak})$, i.e., the marginal increase in the cap equals marginal benefit from peak-time transportation quality, a user's cost accompanied by decreasing congestion equals his or her received benefit and all input choices become socially optimal.

In many cases, it would be difficult to measure marginal benefit from peak-time transportation quality, $v'(q^{peak})$. In practice, it would be easier to implement PC regulation, in which the cap is contingent on capacity investment. Therefore, we next consider this kind of PC regulation,

$$tcap + \overline{t}(L, Z) \geq t. \tag{7}$$

where $\overline{t}(L, Z)$ is the variable part of the ceiling price that depends on railroad right-of-way, L, and the non-capital input, Z. We call this method investment-contingent PC regulation. The maximisation of the present discounted value of the railway firm's profits, (1), subject to the production constraint, (3), and investment-contingent PC regulation, (7) yields proposition 5.

Proposition 5

Under investment-contingent PC regulation, the distortion between the non-capital input, Z, and cost-reducing efforts, e, is never eliminated: $\frac{F_e}{F_Z} > \varphi'(e)$. *If* $\overline{t}_L N = R - tN_L$ *and* $\overline{t}_Z N = 1$,

(i) The marginal benefit of the investment in the non-capital input, Z, equals its marginal cost: $v'(q^{peak}) q^{peak\prime}(q) F_Z = 1$,

(ii) The marginal rate of technical substitution between the non-capital input, Z, and railroad right-of-way, L, equals their relative price: $\frac{F_L - qN_L}{F_Z} = R - tN_L$.

The proof is stated in Appendix 1. Investment-contingent PC regulation can remove quality distortions, as long as the marginal investment costs of both inputs are fully recovered through the increase in the ceiling price, i.e., $\overline{t}_L N = R - tN_L$ and $\overline{t}_Z N = 1$. However, it yields distortions in cost-reducing efforts. This feature stems from the fact that investment-contingent PC regulation lowers the investment costs of railroad right-of-way, L, and non-capital, Z, but leaves the cost of cost-reducing efforts, e, unchanged. Suppose that the marginal investment costs are fully recouped by the increase in the allowed price. In this case, if a railway firm invests in L and Z, the allowed price goes up and the investment costs are virtually zero. If it invests in cost-reducing efforts, however, the allowed price remains unchanged. This asymmetry weakens cost-reducing incentives, and leads to the railway firm operating less efficiently.

Investment-contingent PC regulation would be a possible candidate as a regulation system for congested urban railways, because it can ease congestion and can be implemented based on the data on investment costs of each input, which are easier for the regulatory authority to obtain than the data on marginal benefit from peak-time transportation quality. However, the regulator should be aware that investment-contingent PC regulation weakens the railway firms' incentives to operate efficiently.

7 Conclusion

Our analysis stated above shows that simple PC regulation without properly addressing congestion problem would not be suitable for Japanese urban railways that are extremely congested, because it has adverse effects on congestion although it promotes cost reduction. In modified PC regulations, quality-contingent PC regulation is consistent with the relief of congestion and stays free of distortions in cost-reducing efforts. Although investment-contingent PC regulation also eases congestion, it distorts incentives for cost reduction.

Although our model clarifies the effects of ROR, PC, quality-contingent PC, and investment-contingent PC regulations, it is important to show how serious those effects are. For example, how much does social welfare increase or decrease by the regulatory shift from ROR regulation to PC regulation? Or, how much merits arise when the regulator adopts quality-contingent PC regulation or investment-contingent PC regulation? To answer these questions, we need numerical simulations in which actual railway data are used, which are delegated to future researches.

Appendix 1: Proof of Propositions

Proof of Proposition 1

We set up the Lagrangian for the first best case as follows:

$$\Lambda = PV(SW) + \lambda[F(L,Z,e) - qN(t,q^{peak}(q),L)], \tag{A1}$$

where $\lambda \geq 0$.

The first order conditions for (A1) are

$$\frac{\partial\Lambda}{\partial t} = \frac{1}{r}(N + tN_t + R_t d\bar{H} - R_t L) + \frac{R_t(1-d)\bar{H}}{r} - \lambda q N_t = 0, \tag{A2}$$

$$\frac{\partial\Lambda}{\partial q} = \frac{1}{r}(tN_q + R_q d\bar{H} - R_q L) + \frac{R_q(1-d)\bar{H}}{r} - \lambda(N + qN_q) = 0, \tag{A3}$$

$$\frac{\partial\Lambda}{\partial L} = \frac{1}{r}(tN_L - R) + \lambda(F_L - qN_L) = 0, \tag{A4}$$

$$\frac{\partial\Lambda}{\partial Z} = -\frac{1}{r} + \lambda F_Z = 0, \tag{A5}$$

$$\frac{\partial\Lambda}{\partial e} = -\frac{\varphi'(e)}{r} + \lambda F_e = 0, \tag{A6}$$

$$\lambda[F(L,Z,e) - qN(t,q^{peak}(q),L)] = 0.$$

From $R_{q^{peak}} = \frac{v'(q^{peak})}{h} > 0$ and $N_{q^{peak}} = -\frac{Nh_R v'(q^{peak})}{h^2} > 0$, we derive $R_q = \frac{v'(q^{peak})q^{peak\prime}(q)}{h}$ and $N_q = -\frac{Nh_R v'(q^{peak})q^{peak\prime}(q)}{h^2}$. Using $R_t = -\frac{1}{h}$, $N_t = \frac{Nh_R}{h^2}$, $R_q = \frac{v'(q^{peak})q^{peak\prime}(q)}{h}$, and $N_q = -\frac{Nh_R v'(q^{peak})q^{peak\prime}(q)}{h^2}$, from (A2)-(A6) we obtain

$$v'(q^{peak})q^{peak\prime}(q)F_Z = 1, \tag{A7}$$

$$\frac{F_L - qN_L}{F_Z} = R - tN_L, \tag{A8}$$

$$\frac{F_e}{F_Z} = \varphi'(e). \tag{A9}$$

From (A7)-(A9), we obtain Proposition 1.

Q.E.D.

Proof of Proposition 2

The Lagrangian for profit-maximisation under ROR regulation can be formulated as

$$\begin{aligned}\Lambda &= PV(\pi) \\ &+ \lambda\left[\rho\left\{\frac{R(\bar{y}\text{-}t, q^{peak}(q))}{r}(L-\beta d_2\bar{H}) + \beta\bar{V}d_2\bar{H}\right\} - tN(t, q^{peak}(q), L) + Z\right] \\ &+ \gamma[F(L,Z,e) - qN(t, q^{peak}(q), L)],\end{aligned} \tag{A10}$$

where $\lambda \geq 0$ and $\gamma \geq 0$.

The first order conditions for (A10) are

$$\frac{\partial\Lambda}{\partial t} = \frac{1}{r}(N + tN_t + R_t d\bar{H} - R_t L) + \lambda\left\{\frac{\rho R_t}{r}(L-\beta d_2\bar{H}) - (N + tN_t)\right\} - \gamma q N_t = 0, \tag{A11}$$

$$\frac{\partial\Lambda}{\partial q} = \frac{1}{r}(tN_q + R_q d\bar{H} - R_q L) + \lambda\left\{\frac{\rho R_q}{r}(L-\beta d_2\bar{H}) - tN_q\right\} - \gamma(N + qN_q) = 0, \tag{A12}$$

$$\frac{\partial\Lambda}{\partial L} = \frac{1}{r}(tN_L - R) + \lambda(\frac{\rho R}{r} - tN_L) + \gamma(F_L - qN_L) = 0, \tag{A13}$$

$$\frac{\partial\Lambda}{\partial Z} = -\frac{1}{r} + \lambda + \gamma F_Z = 0, \tag{A14}$$

$$\frac{\partial\Lambda}{\partial e} = -\frac{\varphi'(e)}{r} + \gamma F_e = 0, \tag{A15}$$

$$\lambda\left[\rho\left\{\frac{R(\bar{y}\text{-}t, q^{peak}(q))}{r}(L-\beta d_2\bar{H}) + \beta\bar{V}d_2\bar{H}\right\} - tN(t, q^{peak}(q), L) + Z\right] = 0,$$

$$\gamma[F(L,Z,e) - qN(t, q^{peak}(q), L)] = 0.$$

If ROR regulation is binding, i.e., $\lambda > 0$, (A11)-(A15) yields

$$v'(q^{peak})q^{peak\prime}(q)F_Z = 1, \tag{A16}$$

$$\frac{F_L - qN_L}{F_Z} = R - tN_L - \frac{R\lambda}{1 - r\lambda}(\rho - r), \tag{A17}$$

$$\frac{F_e}{F_Z} = \frac{\varphi'(e)}{1-r\lambda} > \varphi'(e), \tag{A18}$$

where $1-r\lambda \geq 0$ from (A14). From (A16)-(A18), we obtain Proposition 2.

Q.E.D.

Proof of Corollary 1

From (A12) and (A13), we obtain

$$1-r\lambda = \frac{(\rho - r)\left\{R_q L(F_L - qN_L) + R(N+qN_q)\right\} + \left(d - \frac{\rho}{r}\beta d_2\right) r R_q \bar{H}(F_L - qN_L)}{\rho R_q (L - \beta d_2 \bar{H})(F_L - qN_L) + \rho R(N+qN_q) - rt(N_q F_L + NN_L)}.$$

If $1-r\lambda = 0$, then $\frac{F_e}{F_Z} = \infty$ from (A18), i.e., the incentives for cost reduction completely disappear. When $\rho = r$ and $\beta = 1$, $1-r\lambda = 0$ if $d - d_2 \equiv d_1 = 0$. When $\rho = r$ and $\beta = 0$, $1-r\lambda = 0$ if $d = 0$.

Q.E.D.

Proof of Proposition 3

We set up the Lagrangian for PC regulation as follows:

$$\begin{aligned}\Lambda = {} & PV(\pi) \\ & + \lambda[\bar{t} - t] \\ & + \gamma[F(L,Z,e) - qN(t, q^{peak}(q), L)],\end{aligned} \tag{A19}$$

where $\lambda \geq 0$ and $\gamma \geq 0$. The first order conditions for (A19) are

$$\frac{\partial \Lambda}{\partial t} = \frac{1}{r}(N + tN_t + R_t d\bar{H} - R_t L) - \lambda - \gamma q N_t = 0, \tag{A20}$$

$$\frac{\partial \Lambda}{\partial q} = \frac{1}{r}(tN_q + R_q d\bar{H} - R_q L) - \gamma(N + qN_q) = 0, \tag{A21}$$

$$\frac{\partial \Lambda}{\partial L} = \frac{1}{r}(tN_L - R) + \gamma(F_L - qN_L) = 0, \tag{A22}$$

$$\frac{\partial \Lambda}{\partial Z} = -\frac{1}{r} + \gamma F_Z = 0, \tag{A23}$$

$$\frac{\partial \Lambda}{\partial e} = -\frac{\varphi'(e)}{r} + \gamma F_e = 0\,, \tag{A24}$$

$$\lambda[\bar{t} - t] = 0\,,$$

$$\gamma[F(L, Z, e) - qN(t, q^{peak}(q), L)] = 0\,.$$

If PC regulation is binding, i.e., $\lambda > 0$, from (A20)-(A24) we obtain

$$v'(q^{peak})q^{peak\prime}(q)F_Z = \frac{1}{1 - \frac{r\lambda}{N}} > 1\,, \tag{A25}$$

$$\frac{F_L - qN_L}{F_Z} = R - tN_L\,, \tag{A26}$$

$$\frac{F_e}{F_Z} = \varphi'(e)\,. \tag{A27}$$

From (A25)-(A27), we obtain Proposition 3.

Q.E.D.

Proof of Proposition 4

We set up the Lagrangian for quality-contingent PC regulation as follows:

$$\begin{aligned} &\Lambda = PV(\pi) \\ &+\lambda[tcap + \overline{t}\,(q^{peak}(q)) - t] \\ &+\gamma[F(L, Z, e) - qN(t, q^{peak}(q), L)] \end{aligned} \tag{A28}$$

where $\lambda \geq 0$ and $\gamma \geq 0$. The first order conditions for (A28) are

$$\frac{\partial \Lambda}{\partial t} = \frac{1}{r}(N + tN_t + R_t d\overline{H} - R_t L) - \lambda - \gamma qN_t = 0, \tag{A29}$$

$$\frac{\partial \Lambda}{\partial q} = \frac{1}{r}(tN_q + R_q d\overline{H} - R_q L) + \lambda \overline{t}'(q^{peak})q^{peak\prime}(q) - \gamma(N + qN_q) = 0, \tag{A30}$$

$$\frac{\partial \Lambda}{\partial L} = \frac{1}{r}(tN_L - R) + \gamma(F_L - qN_L) = 0, \tag{A31}$$

$$\frac{\partial \Lambda}{\partial Z} = -\frac{1}{r} + \gamma F_Z = 0, \tag{A32}$$

$$\frac{\partial \Lambda}{\partial e} = -\frac{\varphi'(e)}{r} + \gamma F_e = 0\,, \tag{A33}$$

$$\lambda[tcap + \overline{t}\,(q^{peak}(q)) - t] = 0\,,$$

$$\gamma[F(L,Z,e) - qN(t, q^{peak}(q), L)] = 0\,.$$

From (A29)-(A33), we obtain

$$v'(q^{peak})q^{peak\prime}(q)F_Z = 1 - \frac{r\lambda q^{peak\prime}(q)F_Z}{N}\left\{\overline{t}'(q^{peak}) - v'(q^{peak})\right\}, \tag{A34}$$

$$\frac{F_L - qN_L}{F_Z} = R - tN_L\,, \tag{A35}$$

$$\frac{F_e}{F_Z} = \varphi'(e)\,. \tag{A36}$$

From (A34)-(A36), we obtain Proposition 4.

Q.E.D.

Proof of Proposition 5

We set up the Lagrangian for investment-contingent PC regulation as follows:

$$\begin{aligned} \Lambda = {} & PV(\pi) \\ & + \lambda[tcap + \overline{t}\,(L,Z) - t] \\ & + \gamma[F(L,Z,e) - qN(t, q^{peak}(q), L)], \end{aligned} \tag{A37}$$

where $\lambda \geq 0$ and $\gamma \geq 0$. The first order conditions for (A37) are

$$\frac{\partial \Lambda}{\partial t} = \frac{1}{r}(N + tN_t + R_t d\overline{H} - R_t L) - \lambda - \gamma q N_t = 0, \tag{A38}$$

$$\frac{\partial \Lambda}{\partial q} = \frac{1}{r}(tN_q + R_q d\overline{H} - R_q L) - \gamma(N + qN_q) = 0, \tag{A39}$$

$$\frac{\partial \Lambda}{\partial L} = \frac{1}{r}(tN_L - R) + \lambda \overline{t}_L + \gamma(F_L - qN_L) = 0, \tag{A40}$$

$$\frac{\partial\Lambda}{\partial Z} = -\frac{1}{r} + \lambda\bar{t}_Z + \gamma F_Z = 0, \tag{A41}$$

$$\frac{\partial\Lambda}{\partial e} = -\frac{\varphi'(e)}{r} + \gamma F_e = 0\,, \tag{A42}$$

$$\lambda[tcap + \bar{t}\,(L,Z) - t] = 0\,,$$

$$\gamma[F(L,Z,e) - qN(t, q^{peak}(q), L)] = 0\,,$$

Rewriting (A38)-(A42) yields

$$v'(q^{peak})q^{peak\prime}(q)F_Z = \frac{1 - r\lambda\bar{t}_Z}{1 - r\lambda\dfrac{1}{N}}\,, \tag{A43}$$

$$\frac{F_L - qN_L}{F_Z} = R - tN_L - \frac{r\lambda\{\bar{t}_L - \bar{t}_Z(R - tN_L)\}}{1 - r\lambda\bar{t}_Z}\,, \tag{A44}$$

$$\frac{F_e}{F_Z} = \frac{\varphi'(e)}{1 - r\lambda\bar{t}_Z}\,. \tag{A45}$$

As long as investment-contingent PC regulation is binding, i.e., $\lambda > 0$, $\dfrac{F_e}{F_Z} > \varphi'(e)$ from (A45). If $\bar{t}_Z N = 1$ and $\bar{t}_L N = R - tN_L$, i.e., the marginal investment costs of L and Z are fully recovered through the increase in $\bar{t}\,(L,Z)$, then (A43) and (A44) respectively are

$$v'(q^{peak})q^{peak\prime}(q)F_Z = 1\,,\ \frac{F_L - qN_L}{F_Z} = R - tN_L\,.$$

Q.E.D.

References

Armstrong, M., S. Cowan, and J. Vickers, (1994), "*Regulatory Reform*," The MIT Press.

Averch, H. and L. Johnson, (1962), "Behaviour of the Firm Under Regulatory Constraint," *American Economic Review 52*, 1052-1069.

Kanemoto, Y., (1980), "*Theories of urban externalities*," North-Holland.

Kanemoto, Y. and K. Kiyono, (1993), "Investment, Pricing, and Regulation in Urban Transportation and Spatial Development," in H. Kohno and P. Nijkamp (eds.) *Potentials and Bottlenecks of Spatial Economics Development*, 30-44, Springer-Verlag.

Kanemoto, Y. and K. Kiyono, (1995), "Regulation of Commuter Railways and Spatial Development," *Regional Science and Urban Economics 25*, 377-394.

Kidokoro, Y., (1998), "Rate of Return Regulation and Rate Base Valuation," *Regional Science and Urban Economics 28*, 629-654.

Littlechild, S. C., (1983), "*Regulation of British Telecommunications' Profitability*," Report to the Secretary of State, Department of Industry, London.

Spence, A. M., (1975), "Monopoly, quality and regulation," *Bell Journal of Economics 6*, 417-429.

Train, K. E., (1991), "*Optimal Regulation*," The MIT Press.

Varian, H. R., (1992), "*Microeconomic Analysis*," Norton.