SERIE DOCUMENTOS DE TRABAJO

FACULTAD ECONOMÍA Y NEGOCIOS
UNIVERSIDAD DE CHILE

DEPARTAMENTO DE ECONOMÍA

SDT 336

MONOPOLY, SUBSIDIES AND THE MOHRING EFFECT: A SYNTHESIS AND AN EXTENSION

Autor: Andrés Gómez-Lobo

Santiago, Abril de 2011

Monopoly, subsidies and the Mohring effect: A synthesis and an extension[1]

Andrés Gómez-Lobo

Department of Economics

University of Chile

Abstract

This paper discusses the recent literature concerning the Mohring effect and the need to subsidize public transport in order to provide optimal frequencies when there is a monopoly provider. We show that all of the results of this literature are special cases of Spence (1975), albeit with a small adjustment in order to take into account the cost structure of frequency provision in the case of public transport. Although in theory there are cases when a monopolist will offer optimal or above optimal levels of frequency without requiring subsidies, we argue that this result is not very relevant from a public policy perspective. Public transport is rarely provided by an unregulated monopolist. Rather, these services are usually provided either by an exclusive operator under regulated fares or by a group of competing operators, with or without fare regulation. We show that in the first case frequency will always be below social optimal level and in the second case frequency may be overprovided under certain conditions particularly if fares are high. The implications of these results are discussed in the conclusions.

Keywords: Mohring effect, public transport, subsidies, Monopoly, quality provision

JEL Classification: L12, L91, R48

[1] I would like to thank Anna Matus and conference participants for useful comments received during the "Primer Encuentro de Economía de Transportes", FEDEA, Madrid, March 14th, 2011.

1. Introduction

Significant sums of money are spent each year around the world in subsidizing public transport. Are these subsidies justified? There are two main efficiency arguments for subsidizing public transport.[2]

One is the Mohring effect (Mohring, 1972; Jansson, 1979, 1993). This states that as demand for public transport increases, optimal frequencies also increase, diminishing waiting times (when services are not scheduled) or schedule delays costs (when there are service schedules) for all users. Thus, additional demand generates a positive externality on existing users and social marginal costs, which include users waiting time, are below private marginal costs, calling for a subsidy in the first-best pricing solution.

The second efficiency case for transport subsidies is the "second-best" argument related to the need to discourage private transport which is usually underpriced compared to its social cost. If public and private transport trips are substitutes, then a subsidy for public transport may be efficient if it reduces the negative externality related to private automobile usage.[3]

There is a current debate as to whether the Mohring effect justifies subsidizing public transport. This discussion was started by a provocative paper by Van Reeven (2008). It was later rebutted by Basso and Jara-Diaz (2010) and Savage and Small (2010).

What Van Reevan (2008) argues is that since service frequency is a quality variable (affecting willingness to pay for public transport trips) a monopolist will take this into account when setting prices and may well set frequency above the social optimum in order to raise prices and increase profits. He then concludes that since most public transport is provided by a monopolist, subsidies are not warranted, at least based on the Mohring effect.

[2] The distributive or social case for public transport subsidies is discussed in Estupiñan, Gómez-Lobo, Muñoz-Raskin and Serebrisky (2009).

[3] Small and Verhoef (2007; chapter 4) present an excellent exposition and discussion of the issues related to these two efficiency justifications for subsidizing public transport. Empirically, these arguments can be very relevant and justify higher subsidies than currently provided as shown by Parry and Small (2009) for the case of Los Angeles, Washington D.C. and London.

Basso and Jara-Diaz (2010) show that this result depends critically on the demand assumptions made.[4] A slight modification of Van Reevan's (2008) model, introducing heterogeneity in users' reservation price, results in suboptimal frequency provision by a monopolist operator and subsidies would once again be justified in order to increase frequencies.

Karamychev and Van Reeven (2010) counter by generalizing the demand model of Basso and Jara-Diaz (2010) and show that as users' heterogeneity decreases, at some point a monopolist will provide social optimal frequencies and beyond that point will over-supply frequencies compared to the social optimum.

In this paper I want to show explicitly that all of the above results are just special cases of Spence (1975) and thus Karamychev and Van Reeven (2010) results are not at all surprising. Although this is recognized by Van Reeven (2008), Basso and Jara-Diaz (2010) and Savage and Small (2010), I believe it is well worth to carefully show this in a summary paper such as this one.[5] In addition, in so doing I show that Spence's (1975) results depend on a particular assumption regarding the cost of providing quality. This assumption is not reasonable in the context of public transport. All of the above papers model the cost of frequency provision in such a way as to introduce a new effect in Spence's (1975) comparison between the social optimal and the monopolist's frequency provision. Thus, this comparison depends both on demand characteristics (as emphasized by Spence (1975)) as well as cost structure considerations.

More important, we discuss the relevance of Van Reeven (2008) and Karamychev and Van Reeven (2010) for policy applications. We argue that even if in theory there are cases when a monopolist may provide above social optimal frequencies, this is not very interesting from a policy perspective. It is rare for public transport services to be provided by an unregulated monopolist.

The most common market structure in developed countries is private concessionaires or public companies operating routes under exclusivity but with regulated fares, sometimes

[4] Savage and Small (2010) make the same point regarding Van Reeven (2008).

[5] Frankena, (1983) is another example of the application of Spence's results, although emphasizing the case of a monopoly firm that maximizes ridership or miles offered (another way to describe frequency) subject to a budget constraint.

after a tendering process. We can call this the “regulated monopoly” model. In developing countries (as well as in the United Kingdom outside of London) transport services are usually provided under competitive conditions, with multiple operators vying for passengers. Fares in this case are also sometimes set by a regulator. We will call this the “regulated competition” model.

In both of the above cases, the results discussed above regarding monopoly and quality provision do not apply. Furthermore, it is unlikely that a good policy option for providing optimal frequencies will be to liberalize prices and guarantee unregulated monopoly provision in public transport services. Besides the political difficulties and the deadweight loss from monopoly pricing that this policy recommendation would generate, the Mohring effect is not the only justification for subsidizing public transport. There is also a “second best” argument mentioned above. Therefore, unregulated monopoly will most probably result in prices that are grossly above social optimal levels, discouraging public transport use and exacerbating the welfare losses from externalities of private automobile use.

Therefore, if unregulated monopoly provision is not a viable policy alternative, it is relevant to analyze whether frequencies are under or over supplied in the more relevant market structures described above (regulated monopoly and competition). This provides insights as to whether the Mohring effect is still a relevant justification for subsidies.

We find that under a single operator provision, frequencies will always be underprovided compared to the social optimum when fares are fixed. This result was also shown by Spence (1975). Thus, under a regulated monopoly model, subsidies are always justified based on the Mohring effect if fares are set at the first best level. However, under an oligopolistic market structure, frequencies may be over o under provided under certain conditions, particularly if fares are high compared to costs.

This last result is interesting because it may explain why recent reforms in developing countries —that replaced the previous regulated competition model for regulated monopoly model— seem to under-provide frequencies from the perspective of users. In fact, one of the most important complaint of users of Bogota´s well known Transmilenio BRT system and Santiago’s Transantiago reform, are frequency levels.

This suggests that the previous competitive system provided higher frequencies compared to the new tendered concession system. This paper provides one explanation for such a result.

This paper is organized as follows. We first provide a summary of Spence (1975) and show the effects that cost assumptions have on those results. We then show that the results of Van Reeven (1975), Basso and Jara-Diaz (2010) and Karamychev and Van Reeven (2010) easily follow as special cases of Spence´s analysis. We then go on to argue that all this discussion is not very relevant from a policy perspective. Of more relevance is the analysis of the incentives to provide frequency under a regulated monopoly or competition scenario. To this end we develop a model showing that under an oligopoly market structure, frequency may be higher or lower than the social optimum depending on the price level and other conditions. In the conclusions we discuss the implications of this result in light of recent public transport reforms in Latin America as well as present a summary of the results of the paper.

2. Monopoly and quality provision (Spence, 1975)

In this section we summarize Spence's (1975) results concerning the relative supply of quality by a monopolist versus the social optimum. We emphasize the particular cost assumptions made by Spence (1975) and how the results are modified in the case of cost structures more reasonable for frequency provision in the case of public transport.

Assume that the inverse demand for public transport is given by: $p = D(q,b)$, where p is willingness to pay, q is the quantity of rides and b is the number of buses in operation per period of time. Here b determines frequency and thus is the "quality" variable in our model. Owing to the decrease in waiting times when frequency is higher, $\partial p/\partial b > 0$ and, as is standard, $\partial p/\partial q < 0$.

The social optimal level of rides and buses (frequency) will be given by:

$$\max_{q,b} W = \int_0^q D(\theta,b)d\theta - c(q,b) \qquad (1)$$

First order conditions are given by,

$$p = D(q^w, b^w) = \frac{\partial c}{\partial q}(q^w, b^w) \qquad (2)$$

$$\int_0^{q^w} \frac{\partial D}{\partial b}(\theta, b^w) d\theta = \frac{\partial c}{\partial b}(q^w, b^w) \qquad (3)$$

Condition (2) is the normal price equal marginal cost condition of optimal quantity provision for welfare maximization while (3) indicates that frequency should be increased until the addition consumer surplus generated equals the additional costs borne.

A monopolist however, will offer a number of rides and frequency to maximize profits:

$$\max_{q,b} \ \pi = q \cdot D(q, b) - c(q, b) \qquad (4)$$

The first order conditions for this problem are:

$$D(q^\pi, b^\pi) + q^\pi \cdot \frac{\partial D}{\partial q}(q^\pi, b^\pi) = \frac{\partial c}{\partial q}(q^\pi, b^\pi) \qquad (5)$$

$$q^\pi \cdot \frac{\partial D}{\partial b}(q^\pi, b^\pi) = \frac{\partial c}{\partial b}(q^\pi, b^\pi) \qquad (6)$$

Condition (5) is the typical monopolist solution for quantity where marginal income has to be equal to marginal cost, while (6) states that the optimal frequency provision for a monopolist is where the additional income generated equals the additional costs. The additional income from increasing frequency is due to the fact that increasing quality (frequency in our application) allows the monopolist to charge a bit more for the good sold which is determined by the marginal valuation of quality of the marginal consumer. This increase in the price the monopolist can charge is multiplied by all the units sold.

The first question one can ask is whether a monopolist will under or over-provide frequency for a given number of rides? That is, given that the monopolist offers q, is the quality provided above or below the social optimal for that level of supply? This amounts to comparing the social optimal level of quality provision for that quantity of supply with the optimal provision of the monopolist. From conditions (3) and (6), this amounts to comparing:

$$\int_0^q \frac{\partial D}{\partial b}(\theta, b)d\theta \gtreqless q \cdot \frac{\partial D}{\partial b}(q, b)$$

or

$$\frac{\int_0^q \frac{\partial D}{\partial b}(\theta, b)d\theta}{q} \gtreqless \frac{\partial D}{\partial b}(q, b) \qquad (7)$$

The left-hand of (7) is the average valuation of an additional unit of frequency, while the right-hand side is the valuation of frequency by the marginal user ordered according to their willingness to pay for rides. When the average valuation is larger than the marginal valuation —at the given supply level— then the social optimal quality provision is higher than what a monopolist would provide. The converse is also true.

It is also trivial to show that the relative magnitude of the average and marginal valuation will depend on the cross derivative of the inverse demand function.[6] Thus,

$$\frac{\partial^2 D}{\partial b \partial q} > 0 \Rightarrow b^{\pi}(q) > b^{w}(q)$$

$$\frac{\partial^2 D}{\partial b \partial q} < 0 \Rightarrow b^{\pi}(q) < b^{w}(q)$$

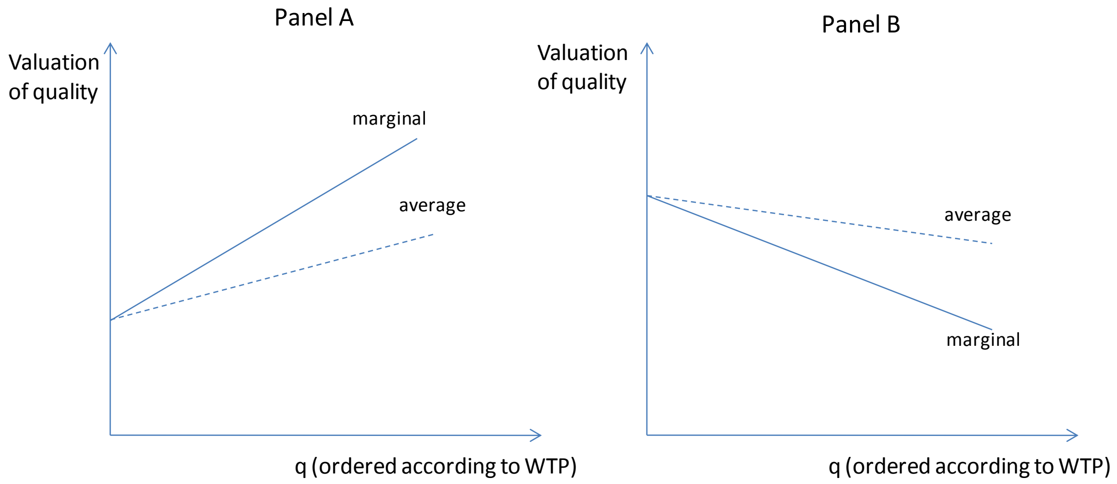
$$\frac{\partial^2 D}{\partial b \partial q} = 0 \Rightarrow b^{\pi}(q) = b^{w}(q)$$

That is, if users are ordered according to their willingness to pay for rides, how the marginal willingness to pay for quality changes along this line will determine whether a monopolist offers more or less quality than the social optimum. Figure 1 of Spence

[6] See Appendix 1.

(1975), shows that if the marginal valuation for quality is increasing, then it will be above the average valuation and the monopolist will offer a higher quality than the social optimum for the same level of output (Panel A). The opposite is true if the marginal valuation of quality is decreasing with output (Panel B). If the cross derivative is cero, then a monopolist will offer the social optimal quality level for the same level of output.

Figure 1: Average versus marginal valuation of quality



However, another question can be posed: whether the frequency provided by a monopolist is above or below the social optimal frequency but evaluated at the social optimal output. In other words, how do $b^{\pi}(q^{\pi})$ and $b^{w}(q^{w})$ compare? This amounts to comparing the frequency provided by a monopolist at the monopoly output level with the social optimal frequency at the social optimal output level.

Here Spence (1975) makes the following cost assumption:

$$c(q, b) = c(b) \cdot q \qquad (8)$$

That is, costs are separable between quality and output and moreover the cost function is linear in output. With this assumption, for a given level of quality, b, the comparison between the monopolist's solution and the social optimum will depend on the relative value of the first order conditions (3) and (6):

$$\int_0^{q^w} \frac{\partial D}{\partial b}(\theta, b) d\theta = c^{'}(b) \cdot q^w \qquad (3')$$

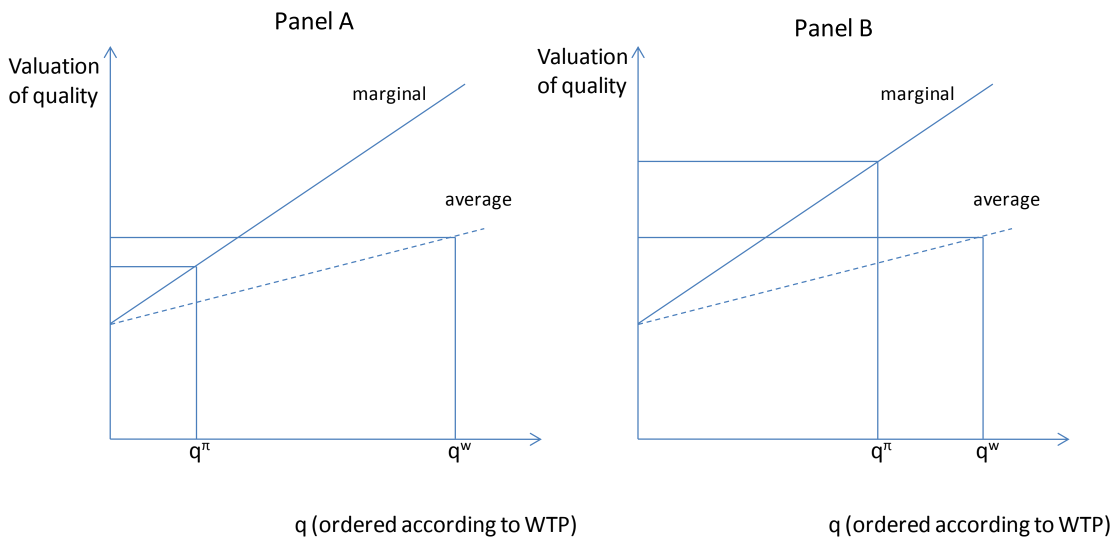
$$q^\pi \cdot \frac{\partial D}{\partial b}(q^\pi, b) = c^{'}(b) \cdot q^\pi \qquad (6')$$

Condition (6') simplifies to $\frac{\partial D}{\partial b}(q^\pi, b) = c^{'}(b)$. Therefore, because the marginal cost of providing quality is the same, the comparison between the social optimal and monopoly solutions amounts to comparing:

$$\frac{\int_0^{q^w} \frac{\partial D}{\partial b}(\theta, b) d\theta}{q^w} \gtreqless \frac{\partial D}{\partial b}(q^\pi, b) \qquad (9)$$

That is, the comparison is between the average valuation of quality but at the social optimal supply level and the marginal valuation of quality at the monopoly supply level. Thus, Spence (1975) concludes that besides the cross derivative of demand between quantity and quality, emphasized above, the elasticity of demand for quantity is also relevant; how q^π compares to q^w is important. If the monopolist severely restricts output then the social optimal quality level may well be above the monopolist's even when the cross derivative of demand is positive (Figure 2, Panel A). However, if demand conditions imply that the monopoly output level is close to the social optimal, then the cross derivative effect dominates (Figure 2, Panel B). Obviously, the signs of the effects are reversed if the cross derivative of demand is negative.

Figure 2: Optimal provision of quality taking into account output level



One important consequence of the above analysis is that if the cross derivative between quality and quantity is cero the monopoly solution coincides with the social optimal solution independently of the quantity supplied in each case. This is due to the cost assumption made. Namely, that there are constant economies of scale with respect to quality provision. Thus, reducing output by the monopolist does not increase the cost of producing quality.

However, in the public transport application it is more common to model cost by the following cost function:

$$c(q,b) = c \cdot b \qquad (10)$$

where c is a constant parameter.[7] This is the assumption made by all recent papers discussing the issue of monopoly and the Mohring effect.

With this cost assumption, conditions (3´) and (6´) become;

$$\int_0^{q^w} \frac{\partial D}{\partial b}(\theta, b) d\theta = c \qquad (3´´)$$

$$q^\pi \cdot \frac{\partial D}{\partial b}(q^\pi, b) = c \qquad (6´´)$$

In this case, the comparison between the social optimal and monopoly solutions depends on:

$$\frac{\int_0^{q^w} \frac{\partial D}{\partial b}(\theta, b) d\theta}{q^w} \gtreqless \frac{\partial D}{\partial b}(q^\pi, b) \cdot \left(\frac{q^\pi}{q^w}\right) \qquad (11)$$

The last term in parenthesis is what I denominate a “scale effect” that is related to the fact that as the monopolist restricts output, the cost of providing an additional unit of quality (frequency in this case) per unit of output increases. Therefore, the monopolist

[7] This assumes there are no capacity constraints on buses and abstracts from the issue of bus capacity choice. If there are cost economies in bus capacity then the scale effect to be described further below would be even stronger making the social optimal frequency provision even more likely to be above the monopoly solution.

reduces quality due to this cost effect in addition to the marginal versus average valuation of quality emphasized by Spence (1975).

Rearranging condition (11) yields:

$$\left(\frac{\frac{\partial D}{\partial b}(q^{\pi},b)}{\frac{\int_0^{q^w}\frac{\partial D}{\partial b}(\theta,b)d\theta}{q^w}}\right)\cdot\left(\frac{q^{\pi}}{q^w}\right)\begin{cases}>1\Rightarrow b^{\pi}>b^{w}\\=1\Rightarrow b^{\pi}=b^{w}\\<1\Rightarrow b^{\pi}<b^{w}\end{cases}\qquad(12)$$

3. Special cases

Now it is easy to see that all of the results in the recent literature can be interpreted as special cases of condition (12). We show this in turn for the different papers.

3.1 Van Reeven (2008)

Van Reeven (2008) assumes that utility is given by:

$$U=v-p-\tau\qquad(13)$$

where v is the reservation utility, p is the fare and τ is waiting time cost (if services are not scheduled) or scheduling cost (if services follow a schedule).

When services are not scheduled but headway is constant and users arrive uniformly at stops, the average waiting time cost is given by:

$$\tau=\frac{t}{2\cdot b}\qquad(14)$$

where t is the value of time. In this model all users are assumed to be homogeneous (same value of t and v), therefore demand is particularly simple. All users (denoted by X) will travel if $v\geq p+\frac{t}{2\cdot b}$ and all will not travel if the converse holds. Therefore, with respect to price demand is absolutely inelastic up to $v-\tau$ and drops to cero above that price level. The monopolist's optimal price level will be where demand is positive and so we can concentrate on this portion of the demand curve. Therefore, the demand structure in this model is quite simple:

$$q(p, b) = X \qquad (15)$$

and the inverse demand function is:

$$p = D(q, b) = v - \frac{t}{2 \cdot b} \ for \ q \leq X \qquad (16)$$

It is straightforward to see that in this case the cross derivative of the inverse demand function is cero:

$$\frac{\partial^2 D}{\partial q \cdot \partial b} = 0 \qquad (17)$$

Therefore, following Spence (1975) the first term in parenthesis of (12) is equal to 1. In addition, since demand is inelastic, then $q^{\pi} = q^{w} = X$, and the last term in (12) is also equal to one. Therefore, it is no surprise that Van Reeven (2008) finds that the frequency a monopolist will offer is the same as the social optimum.

With scheduling it is assumed that users have a preferred departure time $x \in [0, 1)$ where 1 is the time interval relevant for the problem. Bus departures are denoted by $y_i \in [0, 1)$. Frequency is given by the number of departures. Utility of a user is given by:

$$U = v - p - r \cdot \min_{y_i} |x - y_i| \qquad (18)$$

That is, the maximum utility that an individual with a desired departure time of *x* can achieve is the reservation price minus the fare and the minimum scheduling costs, which are determined by the value of time and the minimum scheduling misalignment given the set of departures.[8]

[8] Van Reeven (2008) assume that $r < t$ because waiting at the bus stop is more expensive than waiting at home or work in the case of scheduled services. However, this is not relevant for the results.

In order to obtain the demand function, we start by noting that for bus y_i the marginal consumers (that is, the ones that are just indifferent from travelling on bus y_i or not making a trip) are those positioned on $y_i \mp \frac{(v-p)}{r}$ in the x interval.[9] Thus demand per bus will be:[10]

$$\frac{q(p,b)}{b} = \frac{2\cdot(v-p)}{r} \cdot X \qquad (19)$$

and total demand is thus:

$$q(p,b) = \frac{2\cdot(v-p)}{r} \cdot X \cdot b \leq X \qquad (20)$$

It is easy to verify that the inverse demand function is given by:

$$p = D(q,b) = v - \frac{r\cdot q}{2\cdot X\cdot b} \qquad (21)$$

Thus, the cross derivative is:

$$\frac{\partial^2 p}{\partial q\cdot\partial b} = \frac{r}{2\cdot X\cdot b^2} > 0 \qquad (22)$$

Thus, the first term of (12) is positive and the monopolist should offer a higher frequency than the social optimal level. Moreover, it is easy to verify that for this particular case, the first term of (12) is:

$$\frac{\frac{\partial D}{\partial b}(q^\pi,b)}{\frac{\int_0^{q^w}\frac{\partial D}{\partial b}(\theta,b)d\theta}{q^w}} = 2\cdot\left(\frac{q^\pi}{q^w}\right) \qquad (23)$$

Therefore, condition (12) will be:

9 This assumes that frequency is not so large as to make all potential users' utility strictly positive. However, in this case the monopolist has incentives to reduce frequency and so will never be optimal to offer such high frequency.

10 There seems to be a typo in Van Reeven (2008) as the *X* term is in the denominator of the expression for *d* while the correct expression has this variable in the numerator.

$$\left(\frac{\frac{\partial D}{\partial b}(q^{\pi},b)}{\frac{\int_0^{q^W}\frac{\partial D}{\partial b}(\theta,b)d\theta}{q^W}}\right)\cdot\left(\frac{q^{\pi}}{q^{w}}\right)=2\cdot\left(\frac{q^{\pi}}{q^{w}}\right)^2 \qquad (24)$$

What is the value of this condition? This is a bit tricky since the final output of the monopolist and the social optimum depend also on frequency. However, the monopoly output evaluated at the social optimal frequency level is:

$$q^{\pi}(b^{w})=v\cdot X\cdot\sqrt{\frac{X}{4\cdot r\cdot c}} \qquad (25)$$

For the Monopolist to produce at all, $v\geq\sqrt{\frac{2\cdot r\cdot c}{X}}$, so we can concentrate on this case,

$$q^{\pi}(b^{w})\geq\sqrt{\frac{2\cdot r\cdot c}{X}}\cdot X\cdot\sqrt{\frac{X}{4\cdot r\cdot c}}=\frac{X}{\sqrt{2}} \qquad (26)$$

Therefore, evaluated at the social optimal frequency level:[11]

$$\frac{q^{\pi}}{q^{w}}(b^{w})\geq\frac{X}{\sqrt{2}}\cdot\frac{1}{X}=\frac{1}{\sqrt{2}} \qquad (27)$$

and,

$$\left(\frac{\frac{\partial D}{\partial b}(q^{\pi},b)}{\frac{\int_0^{q^W}\frac{\partial D}{\partial b}(\theta,b)d\theta}{q^W}}\right)\cdot\left(\frac{q^{\pi}}{q^{w}}\right)=2\cdot\left(\frac{q^{\pi}}{q^{w}}\right)^2\geq 1 \qquad (28)$$

and the oversupply result follows.

3.2 Basso and Jara-Diaz (2010)

Basso and Jara-Diaz (2010) modify slightly Van Reeven's first model to introduce some demand reaction to price in order to show how sensitive Van Reeven's (2008) results

[11] The social optimal output level is *X*.

are to model specifications. They then criticize Van Reeven's (2008) sweeping generalization that the Mohring effect never justifies subsidies.

They assume that the users are not homogenous and instead reservation utilities, *v*, are uniformly distributed in the interval $[\underline{v}, \bar{v}]$. In this case, demand function is:

$$q(p,b) = \frac{X}{(\overline{v}-\underline{v})} \cdot \left(\overline{v} - p - \frac{t}{2 \cdot b}\right) \qquad (29)$$

And the inverse demand function is:

$$p = D(q,b) = \overline{v} - \frac{(\overline{v}-\underline{v})}{X} \cdot q - \frac{t}{2 \cdot b} \qquad (30)$$

It can be easily verified that in this case the cross derivative is also cero, since

$$\frac{\partial p}{\partial q} = \frac{(\overline{v}-\underline{v})}{X} \qquad (31)$$

is independent of frequency *b*. Therefore, in this model, the first term of (12) is equal to one. However, since in this model the demand elasticity is different from cero and the monopolist would charge a higher price than the social optimum, $q^{\pi} < q^{w} = X$. Thus, the second term of (12) is smaller than one implying that the monopolist offers a lower frequency than the social optimal equilibrium. This is precisely what Basso and Jara-Diaz (2010) show. However, this effect is due to the cost assumptions noted above and not because of the demand characteristics as emphasized by Spence (1975). In fact, in Spence's model the demand structure (19) would imply social optimal quality provision by a monopolist.

3.3 Karamychev and Van Reeven (2010)

As a reaction to Savage and Small (2010) and Basso and Jara-Diaz (2010), Karamychev and Van Reeven (2010) develop a model where the distribution of reservation prices is:

$$F(v) = \left(\frac{v}{\bar{v}}\right)^{k} \text{ where } k \in [1, \infty)$$

For $k = 1$ the distribution is uniform as in Basso and Jara-Diaz (2010) but as k grows, heterogeneity decreases until in the limit all reservation prices are the same and there is no heterogeneity.

Using this model Karamychev and Van Reeven (2010) show that there is a critical level of heterogeneity such that for all value of k implying less heterogeneity, the result of excess frequency by a monopolist applies. While for k implying more heterogeneity the under supply result applies. They then argue that Basso and Jara-Diaz (2010) and Savage and Small (2010) assertion that heterogeneity in reservation prices is enough to obtain the under-supply result is wrong since in their model they obtain oversupply of frequency for positive levels of heterogeneity. They also argue that "it is quite reasonable to assume, in a first approximation, that consumer´s reservation prices are concentrated around the entry costs of car ownership and usage" (page 382).

As to the second point, namely that one would expect low user heterogeneity in valuations, this is an empirical question. Karamychev and Van Reeven (2010) argue that since the reservation price depends on the price of entry into private transport which should be similar for different individuals then reservation prices will be similar across the population. However, heterogeneity in reservation prices can arise for many different reasons, including differences in access costs to public transport, residential location, demographic characteristics of the household, and just plain tastes (some people may like using public transport while other do not). Even more important is the fact that heterogeneity will not only be limited to the reservation price, but due to differences in income levels, people will also differ as to the marginal utility of income and value of time. The effects of heterogeneity in these variables for the results are analyzed in the next section. Thus, I conjecture that heterogeneity should be the norm rather than the exception in the valuation of public transport, as well as for most other goods at a microeconomic level.

As for the first point, that heterogeneity in reservation prices will not always generate the under supply result, a clarification is in order. Basso and Jara-Díaz (2010) present a model without schedules. In this context, heterogeneity in reservation prices will in fact always generate the under supply result even under the assumptions of Karamychev and

Van Reeven (2010). To see this note that the demand for public transport in this case would be:[12]

$$q(p,b) = 1 - F\left(p + \frac{t}{2 \cdot b}\right) \qquad (32)$$

Thus, the inverse demand function is:

$$p = F^{-1}(1-q) - \frac{t}{2 \cdot b} \qquad (33)$$

Thus, the cross derivative with respect to quantity and quality is zero. Therefore, the first term of (12) will be one and as long as there is some demand reaction to price, the second term will be smaller than one. As a result, a monopolist will always undersupply frequencies except in the limiting case where k tends to infinity and there is no longer any heterogeneity in reservation prices.

What Karamychev and Van Reeven (2010) require in order to obtain their result is to use the scheduling model of Van Reeven (2008). We already showed that this model implies oversupply by a monopolist. However, heterogeneity in reservation prices generates the opposite effect of undersupply as in Basso and Jara-Diaz (2010). Thus it is no surprise that for scenarios with high reservation price heterogeneity, the undersupply effect dominates, while for low levels of heterogeneity the oversupply effect dominates.

4. Relevance

In sum, in a monopoly situation, subsidies may not be required to provide optimal (or above optimal) frequencies. A necessary condition for this result to hold is that the marginal valuation of frequencies be increasing as we move down the demand curve and that the price elasticity of demand is low. How likely is it that these conditions will hold in a public transport situation?

As for the marginal valuation of frequencies, I am not aware of empirical studies that try

[12] Without loss of generality we follow Karamychev and Van Reeven (2010) and assume $\bar{v} = 1$. We also assume $X = 1$.

to measure how this valuation changes along the demand curve and, anyhow, any such result will probably be driven by the functional form assumptions. However, one can discuss the probable factors that may be involved. For example, assume a slight generalization of the utility specification presented above:

$$U = v - \lambda \cdot p - t \cdot f(b) \qquad (13')$$

where λ is now the marginal utility of income and *f(b)* is a decreasing function of frequency. We know that heterogeneity in *v* among the populations in a model without schedules leads to under provision of frequency by a monopolist. However, λ and *t* are also likely to vary among the user population, particularly due to differences in income. Heterogeneity in *t* is likely to increase frequency supply by a monopolist, just as in the model with schedules in Van Reeven (2008) albeit for another reason. For a given *v* and λ, demand for public transport will be given by: [13]

$$q(p,b) = G\left(\frac{v-\lambda \cdot p}{f(b)}\right) \qquad (34)$$

Where *G(.)* is the cumulative distribution function of *t* in the population. It is easy to verify that the cross derivative of the inverse demand function in this case is:

$$\frac{\partial^2 p}{\partial q \cdot \partial b} = -\frac{\partial G^{-1}}{\partial q} \cdot \frac{f'(b)}{\lambda} > 0 \qquad (35)$$

The intuition is that in this case the marginal user (the one that is indifferent between making or not making the trip) is one with a high value of time, higher than for the average user. Therefore, a monopolist will tend to offer a higher quality just as in the scheduling model of Van Reeven (2008), unless the scale effect more than compensates for this effect.

However, heterogeneity in the marginal utility of income goes the other way.[14] To see this, note that for a given value of time, the demand for transport when there is

[13] For ease of exposition we normalize *X* to 1 in what follows.
[14] Heterogeneity in the marginal utility of income is equivalent to having heterogeneity in reservation utility and in the value of time simultaneously.

heterogeneity in λ is:

$$q(p,b) = H\left(\frac{v - t\cdot f(b)}{p}\right) \qquad (36)$$

where *H(.)* is the cumulative distribution function of λ in the population. It is easy to verify that the cross derivative of the inverse demand function in this case is:

$$\frac{\partial^2 p}{\partial q \cdot \partial b} = \frac{\partial H^{-1}}{\partial q} \cdot \frac{t \cdot f'(b)}{(H^{-1})^2} < 0 \qquad (37)$$

Thus, in this case a monopolist will certainly offer a lower frequency than the social optimum. The intuition is that the marginal consumer is someone with a relatively high marginal utility of income, and thus a low relative valuation of time costs and thus frequency.

To complicate matters, it is reasonable to assume that the value of time and the marginal utility of income are negatively correlated in the population, since the first variable increase with income while the second tends to decrease with income. Therefore, whether a monopolist will offer higher or lower frequency than the social optimum is an empirical matter and nothing can be said in general.

However, even in the case where demand conditions are such that a monopolist would oversupply frequencies and therefore there is no subsidy justification based on the Mohring effect, how relevant would this result be from a public policy perspective? I think it would be very slight.

Public transport is rarely offered by an unregulated monopolist. Van Reeven states that monopoly is "…the dominant organization form in public transport provision" (Van Reeven (2008), pg. 350). However, this confuses the number of suppliers with true monopoly provision, which in addition to exclusivity in supply requires unregulated prices. In most cases in developed countries (with the exception of cities in England and Wales outside of London), public transport services enjoy exclusivity. That is, entry is usually not allowed by competing bus companies. However, in most cases the exclusive provider is not free to set fares and thus cannot be considered a true monopoly market structure. In some cases, such as London, services are franchised to private companies through a competitive process but fares are set by the authorities and payment to bus operators are made according to their contract conditions.[15]

Thus, the dominant market structure in developed countries is exclusive provision but with regulated fares, or what I call a "regulated monopoly" structure.[16] When fares are not regulated there is usually free entry into the market and competition among suppliers. This is the dominant structure in developing countries, although in many cases fares are also set by the authorities. This is what I call the "regulated competition" model, or "unregulated competition" model if fares are not set by a regulator. This is also the case of public transport provision is cities outside of London in England and Wales.

In either of the above cases, the results of Van Reeven (2008) and Karamychev and Van Reeven (2010) do not apply, even in the case were and unregulated monopoly would supply the social optimum frequency levels or more.[17] Given this, would it be a viable or recommendable policy option to liberalize fares and create legal unregulated monopolies in order to provide public transport services without subsidies? I believe not for several reasons.

First, liberalizing fares will create a deadweight loss due to monopoly pricing. Although this deadweight loss may be small if public transport demand is inelastic —as it often is in developing countries— one can also add the monopoly profits as a deadweight loss due to the probable rent seeking behavior that a legal monopoly would generate. Rent seeking behavior could be avoided if the provider is chosen through a tendering process to the bidder willing to pay the highest sum to the authorities. However, even in this last case, it would be politically difficult to explain to users why the tendering mechanism is not based on the lowest fare rather the highest transfer to the authorities.

---

[15] In the case of London, bus operators are paid a fixed sum annually to provide an established number of services and thus their income does not even depend on the number of riders. In France, contracts usually allow for a small portion of income to be related to the number of passengers, but in general can also be considered gross cost contracts.

[16] When the supplier is chosen through a competitive process this is just a "competition for the market" franchise as advocated by Chadwick (1859) almost 150 years ago.

[17] The only case that I am aware of where these results may apply is in the case of shared taxis in Santiago, Chile. They have a bizarre regulation whereby taxi companies have concession contracts that guarantee exclusivity in their service area but are free to set their own fares.

Second, there are other justifications for subsidizing public transport besides the Mohring effect. As mentioned in the introduction, the "second best argument", based on the externalities generated by private transport, can also be an important justification for subsidies. High fares will discourage public transport and encourage private transport use, which creates a social loss if private transport does not pay for the social costs generated. Van Reeven (2008) discounts this possibility by stating that the cross elasticity between car usage and public transport is very low citing two references. However, other sources imply higher cross elasticities. In particular, Perry and Small (2009) show that, based on their references, the cross elasticity is sufficiently high to justify considerable subsidies in the case of Los Angeles, Washington D.C. and London.

Therefore, if unregulated monopoly provision without subsidies is an unlikely policy option, what is frequency provision in the other two market structures discussed above? Furthermore, are subsidies justified in these contexts? In the next two sections we address these questions for the regulated monopoly and competition model.

5. Monopolist with fixed fares

The case of a monopoly provider with a regulated fare is easy to address and was already answered by Spence (1975) in footnote 5 of his article. Social welfare is given by the sum of consumer and producer surplus. If p is fixed, then the social welfare will be maximized with respect to frequency when:

$$\frac{\partial W}{\partial b}(p,b) = \int_p^\infty q_b(\theta,b)d\theta + \frac{\partial \pi}{\partial b} = 0 \qquad (38)$$

where *q(p,b)* is the demand function. The monopolist will offer frequency until $\frac{\partial \pi}{\partial b} = 0$, however we can see that evaluated at this level of frequency $\frac{\partial W}{\partial b} > 0$ and the social optimal frequency level is higher than the monopolists. The intuition for this result is that under a fixed fare, if a private company has incentives to provide more quality then it will always be socially beneficial to provide this additional quality and then some.

Therefore, under the most common market structure of public transport provision in developed countries, frequency will always be underprovided unless other regulations (minimum frequency regulations) are established. If in addition, fares are set at first-best levels then subsidies will be required to induce a private provider to supply optimal frequencies. This can be seen in our model where there are no costs associated with rides and thus the first best fare level is $p = 0$. In this case, subsidies must equal c times the optimal number of buses in service.

6. Oligopolistic model with fixed fares

What about frequencies in a competitive context? We first develop a model with fixed fares. We then discuss what the results would be if fares as well as frequencies are set by a competing set of operators.

Assume there are n operators that compete for passengers with fares set at $\bar{p}$. We assume that the demand of each operator is proportional to the number of buses it has in operation relative to the total number of buses in operation. Thus, each operator's profits are given by:

$$\pi_i = \bar{p} \cdot \frac{b_i}{\sum_j b_j} \cdot q\left(\bar{p}, \sum_j b_j\right) - c(b_i) \quad \forall\, i \in \{1, n\} \qquad (39)$$

We assume for simplicity in the above formulation that there are no costs associated with additional rides, except through the provision of buses.[18]

The first order condition for a maximum is:

$$\frac{\partial \pi_i}{\partial b_i} = \bar{p} \cdot \frac{\left(\sum_j b_j - b_i\right)}{\left(\sum_j b_j\right)^2} \cdot q\left(\bar{p}, \sum_j b_j\right) + \bar{p} \cdot \frac{b_i}{\sum_j b_j} \cdot \frac{\partial q}{\partial b}\left(\bar{p}, \sum_j b_j\right) - \frac{\partial c}{\partial b}(b_i) = 0 \qquad (40)$$

Assuming a symmetric Nash Equilibrium in frequency supply, equal to b, the first order condition for each operator in equilibrium would be:

$$\bar{p} \cdot \frac{(n-1)\cdot b}{(n \cdot b)^2} \cdot q(\bar{p}, n \cdot b) + \bar{p} \cdot \frac{1}{n} \cdot \frac{\partial q}{\partial b}(\bar{p}, n \cdot b) - \frac{\partial c}{\partial b}(b) = 0 \qquad (41)$$

[18] We could add a restriction that demand per bus has to be below bus capacity. However, this would not add much intuition to the model.

It is straightforward to verify that if $n = 1$, then the above condition collapses to the first order condition for a monopoly that faces a fixed price $\bar{p}$. That is, $b(\bar{p}) = b^{\pi}(\bar{p})$ in that case.

When $n > 1$, the above equilibrium can be expressed as:

$$\bar{p} \cdot \frac{(n-1) \cdot b}{n \cdot (b)^2} \cdot q(\bar{p}, n \cdot b) - (n-1) \cdot \frac{\partial c}{\partial b}(b) + \bar{p} \cdot \frac{\partial q}{\partial b}(\bar{p}, n \cdot b) - \frac{\partial c}{\partial b}(b) = 0 \qquad (42)$$

or

$$(n-1) \cdot \left(\bar{p} \cdot \frac{1}{n \cdot b} \cdot q(\bar{p}, n \cdot b) - \frac{\partial c}{\partial b}(b)\right) + \bar{p} \cdot \frac{\partial q}{\partial b}(\bar{p}, n \cdot b) - \frac{\partial c}{\partial b}(b) = 0 \qquad (43)$$

When the marginal cost of providing frequency is constant and equal to c and $b = \frac{b^{\pi}}{n}$, the last two terms sum to cero, since these are identical to the first order condition for a maximum of a monopolist.[19] That is, by definition:

$$\bar{p} \cdot \frac{\partial q}{\partial b}(\bar{p}, b^{\pi}) - c = 0 \qquad (44)$$

Thus, whether a set of firms offer higher frequency than a monopolist will depend on whether:

$$\bar{p} \cdot \frac{q(\bar{p}, b^{\pi})}{b^{\pi}} - c \geq 0 \qquad (45)$$

But this last condition is the monopolist break even constraint. This first term is the fare multiplied by the average demand per bus and the last term is the cost per bus. If the monopolist operates, this condition must be non-negative. Thus we can make the following proposition:

Proposition 1: when public transport fares are fixed and there are constant returns to

[19] If there are diseconomies of scale in providing frequency, then it would be more efficient to have many small firms rather than a monopolist, and an oligopolistic industry would certainly supply more frequency than a monopolist.

scale in frequency provision, oligopoly firms will supply and equal or higher frequency than a monopolist would.

Whether the oligopoly solution provides higher frequency than the social optimum will depend on the relative value of the first terms of (38) and (43):

$$\int_{\bar{p}}^{\infty} q_b(\theta, b) d\theta \gtreqless (n-1) \cdot \left(\bar{p} \cdot \frac{1}{n \cdot b} \cdot q(\bar{p}, n \cdot b) - c\right) \qquad (46)$$

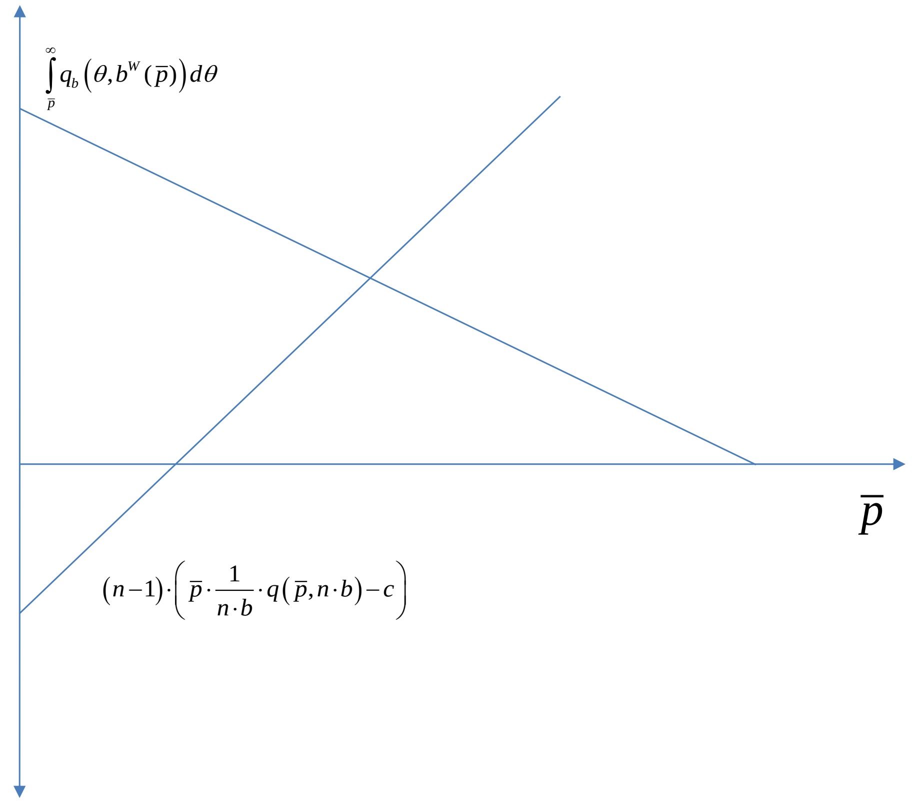
Evaluated at the social optimal frequency provision for each $\bar{p}$, $b^W(\bar{p})$, the left hand side of (46) is decreasing in $\bar{p}$ if q_{bb} and $\frac{db^W}{d\bar{p}}$ are negative, while the right hand side is likely to be increasing in $\bar{p}$, assuming an inelastic demand function for all relevant fare levels and q_b not too big compared to average demand per bus. We also know that if $\bar{p} = 0$, the first best pricing rule in this model (since there are no cost of providing rides), the left hand side is positive while the right hand side is negative, therefore there must be a positive price level where both values coincide. Figure 3 illustrates the argument.

Therefore, we can state the following proposition:

Proposition 2: For a fixed number of firms, constant returns to scale in frequency provision and a fixed fare, there will be a fare level above which a competitive market will offer a frequency above the social optimal level and below which it will offer a frequency below the social optimal level.

The intuition for this result is that competitive firms will increase their supply of frequency in an effort to capture a bigger share of the available demand; as all operators do this the overall frequency level increases. If fares are very high, implying rents, this incentive will be very strong and the aggregate frequency in a competitive equilibrium may be higher than the social optimum. Free entry will make this result even more likely. Evaluated at the social optimal frequency supply, the right hand side of (46) is increasing in *n*.

Figure 3



What if fares are not set by the authorities? Answering this question would require solving a model of fare and frequency determination. In the case of scheduled services, Evans (1987) does this. In the case of unscheduled services, there are reasons to believe that operators will not compete vigorously on fares, due to the search cost of users. Gomez-Lobo (2007) shows that under certain condition this will lead to monopoly pricing, something that has been documented in liberalization experiences around the world.

The point is that it is possible that in competitive urban transport markets fares are sufficiently high (compared to costs) as to induce over supply of frequency in equilibrium. In these cases, subsidies are not required to induce optimal frequencies. In the conclusion we will discuss the relevance of this possibility.

7. Conclusions and discussion

In this paper we presented a synthesis of recent literature concerning whether a monopolist would tend to under or over provide frequency in urban transport. We show that the results are special cases of Spence (1975) duly modified for the cost assumptions made in the literature regarding frequency provision. As is well known, a monopolist may over or under provide frequency depending on demand characteristics. However, ascertaining which is the case in a particular context is very difficult given that the results will depend, among other aspects, on the cross derivative of the inverse demand function with respect to quantity and frequency. In addition, heterogeneity in user's value of time and in the marginal utility of income have opposite effects on frequency provision by a monopolist, so that it is not possible to say in general which is the more likely result.

More importantly in this paper we argue that in spite of the fact that a monopolist may over provide frequency in some contexts this is not a very relevant result from a public policy point of view. Few public transport markets can be characterized as being monopolies. Neither does it seem advisable to create legal monopolies and liberalize fares as a way to provide social optimal frequencies without requiring subsidies.

Most public transport markets can be characterized as either regulated monopolies, where an exclusive operator provides services but with regulated fares, or regulated or unregulated competition. Spence (1975) already showed that in the first case, a private monopoly operator will always provide less than the social optimal level of frequency. Thus, minimum frequency requirements will have to (and usually are) included in franchise contracts in order to guarantee optimal frequencies. In addition, if fares are to be set at first best levels, then subsidies will also be required.

We also show that in the case of a competitive market, frequencies may be over provided if fares are high enough. In this case, subsidies are not requires and in general in developing countries, where competition is the norm, subsidies are not provided to public transit operators. This is also the general case in England and Wales, with the exception of London.

This last result may explain why in recent reforms in Latin America, where the previous competitive system were replaced by regulated monopoly, user's most important complaint has been the low level of frequency of the new system. In Bogotá, Colombia, an 84 kilometer Bus Rapid Transit (BRT) system called Transmilenio was introduced in 2000. This replaced the chaotic and low quality of the services of the prior competitive system with franchised private operators. Operators do not receive a subsidy in this case and are paid according to the conditions bid in the tendering process. Although this system has been very successful in general and has been expanded through the last decade, it is interesting to note that EMBARQ (2009), analyzing survey information from users, conclude that one of the features that need to be improved is frequency levels.

In the case of Santiago, Chile, a very ambitious reform of the public transport system was introduced in 2007 called Transantiago. Unlike the Bogota Transmilenio experience, where only certain corridors of the city were subject to reform leaving the previous competitive system in the rest of the city, in the case of Santiago the competitive system for the whole city was replaced by 14 formal operators under a franchise system. Competition among operators was eliminated overnight.

Initially, Transantiago faced many critical problems; for example, insufficient frequencies and services. In part this was due to lack of subsidies contemplated in the initial reform design. For the same average fare as the old system, the new system had to be financially self sufficient although costs were clearly greater due to new buses, electronic pre-payment cards and better labor conditions for drivers. As Jara-Díaz and Gschwender (2009) show, when financial constraints are binding, planner will tend to increase the size of buses and reduce the fleet.

The initial problems with Transantiago forced the political authorities to introduce a substantial operating subsidy together with an increase in frequencies and fleet size. Although operating conditions are now satisfactory it is interesting to note that the single most frequent complaint by users is still the low frequency of buses, representing 34% of improvement suggestions by users in August-September 2010 (Collect-Gfk, 2010). A possible explanation for this result is that the previous competitive system provided higher frequencies, possibly above social optimal levels. Evidence for this was the observation that in the previous competitive system there seemed to be an excessive number of buses in the streets, mostly empty, particularly during off-peak periods. Although this might have been inefficient, in the eyes of users frequency was better under the old system. As shown in this paper, this can happen if fares are high relative to costs. Something similar may explain user's perceptions in the case of Bogotá

In summary, subsidies will probably still be required to guarantee social optimal frequency levels under first best fare levels. This is particularly so under a regulated monopoly model, the most common market structure in developed countries. The alternative could be to liberalize markets and let competition guarantee high frequency levels. However, this requires fares to be high compared to costs. In addition, as discussed in Gomez-Lobo (2007) there are many other problems related to this type of model in the case of urban transport, particularly the social cost related to accidents as buses compete for passengers in the streets. This has been one of the main motivations for replacing the competitive model for a regulated monopoly model in countries such as Chile and Colombia.

Finally, a word on the incentive effects of subsidies. Van Reeven (2008) seems to have been motivated in part by the inefficiency of transport operators, often public, that have survived thanks to growing public subsidies. However, this X-inefficiency issue related to subsidies need not always occur. If this is the main worry regarding subsidies then a possible alternative is to tender contracts to private operators and pay operators according to the terms of their bid. Any subsidy can then be used to reduce fares from the self financing level. This is the way subsidies operate in systems such as Transantiago. Competition at the tendering stage should guarantee X-efficiency. Any ex-post inefficiency after the tendering stage will not generate higher fares or subsidies, only smaller profits and possibly loses for operators.

References

Basso, L.J. and S.R. Jara-Diaz (2010), 'The Case for Subsidisation of Urban Public Transport and the Mohring Effect', *Journal of Transport Economics and Policy,* 44(3), 365-372.

Chadwick, E. (1859), 'Results of Different Principles of Legislation and Administration in Europe of Competition for the Field as Compared to Competition Within the Field of Service', *Journal of the Royal Statistical Society*, serie A22, pp. 381-420.

Collect-GfK (2010), 'Transantiago: Estudio Tracking Calidad de Servicio N°3, Agosto-Septiembre 2010', Coordinacion available at:
http://www.transantiago.cl/descargas/servicio/Tracking%20(ago-ept%202010%20).pdf

EMBARQ (2009), 'Evaluación Ex-Post Sistema de Transporte Masivo de Bogotá, Fase I y II, Informe 4', report prepared by the Centre for Sustainable Transport of the World Resources Institute for the National Planning Department of the Government of Colombia, November 27, Washington D.C.

Estupiñan, N., A. Gómez-Lobo, R. Muñoz-Raskin y T. Serebrisky (2009), 'Affordability of Public Transport: what do we mean, what can be done?', *Transport Reviews*, Vol. 29(6), pp. 715-739.

Evans, A. (1987), 'A Theoretical Comparison of Competition with other Economic Regimes for Bus Services', *Journal of Transport Economics and Policy*, 13, pp. 268-94.

Frankena, M.W. (1983), 'The Efficiency of Public Transport Objectives and Subsidy Formulas', *Journal of Transport Economics and Policy*, 17, 67-76.

Gómez-Lobo, A. (2007), 'Why Competition Does Not Work in Urban Bus Markets: Some New Wheels for Some Old Ideas', *Journal of Transport Economics and Policy*, Vol. 41, part 2, May, pp. 283-308.

Jansson, J.O. (1979), 'Marginal Cost Pricing of Scheduled Transport Services', *Journal of Transport Economics and Policy*, 13, 268-94.

Jansson, K. (1993), 'Optimal Public Transport Price and Service Frequency', *Journal of Transport Economics and Policy*, 27, 33-50.

Jara-Díaz, S.R. and A. Gschwender (2009), 'The Effect of Financial Constraints on the Optimal Design of Public Transport Services', *Transportation*, 36(1), 65-75.

Karamychev, V. and Van Reeven (2010), 'Oversupply or Undersupply in a Public Transport Monopoly? A rejoinder and Generalization', *Journal of Transport Economics and Policy,* 44(3), 381-89.

Mohring, H. (1972), 'Optimization and Scale Economies in Urban Bus Transportation', *American Economic Review*, 62, 591-604.

Parry, I.W.H and K.A. Small (2009), 'Should Urban Transit Subsidies be reduced?', *American Economic Review*, 99, 700-24.

Savage, I. and K.A. Small (2010), 'A comment on 'Subsidisation of Urban Public Transport and the Mohring Effect'', *Journal of Transport Economics and Policy*, 44(3), 373-80.

Small, K.A. and E.T. Verhoef (2007), *The Economics of Urban Transportion*, Routledge.

Spence, A.M. (1975), 'Monopoly, quality and regulation', *The Bell Journal of Economics*, 6, 417-29.

Van Reeven, P. (2008), 'Subsidisation of Urban Public Transport and the Mohring Effect', *Journal of Transport Economics and Policy*, 42(2), 349-59.

Appendix 1

From (3) and (6) it can be seen that for the same level of production and quality, the marginal costs will be identical in both first order conditions. Therefore, the social optimal quality provision is above the monopolist if:

$$\int_0^q \frac{\partial D}{\partial b}(\theta, b^\pi)d\theta - q \cdot \frac{\partial D}{\partial b}(q, b^\pi) > 0$$

and the vice versa. Now, integrating the first term by parts yields:

$$q \cdot \frac{\partial D}{\partial b}(q, b^\pi) - \int_0^q \theta \cdot \frac{\partial^2 D}{\partial b \partial q}(\theta, b^\pi)d\theta - q \cdot \frac{\partial D}{\partial b}(q, b^\pi) > 0$$

or

$$-\int_0^q \theta \cdot \frac{\partial^2 D}{\partial b \partial q}(\theta, b^\pi)d\theta > 0$$

which will be true whenever the cross derivative is negative.

Appendix 2

As a first step we first show the sign of $\frac{\delta b^W}{\delta \overline{p}}$, that is how the social optimal frequency level changes with the fixed fare. Social welfare is given by:

$$W(\bar{p}, b) = \int_{\bar{p}}^{\infty} q(\theta, b)d\theta + \pi(\bar{p}, b) = 0$$

The first order condition for a maximum is:

$$\frac{\partial W}{\partial b}(\bar{p}, b^W) = \int_{\bar{p}}^{\infty} q_b(\theta, b^W) d\theta + \frac{\partial \pi}{\partial b}(\bar{p}, b^W) = 0$$

or,

$$\frac{\partial W}{\partial b}(\bar{p}, b^W) = \int_{\bar{p}}^{\infty} q_b(\theta, b^W) d\theta + \bar{p} \cdot q_b(\bar{p}, b^W) - c = 0$$

Using the implicit function theorem:

$$\frac{db^W}{d\overline{p}} = \frac{q_b - \overline{p} \cdot q_{bp}}{\int_{\overline{p}}^{\infty} q_{bb} d\theta + q_{bb}}$$

Under the reasonable assumptions that $q_b > 0$ and $q_{bb} < 0$, that is, higher frequency increases demand but at a decreasing rate, the optimal frequency level will be decreasing with respect to the fare level if $q_{pb} \leq 0$ or $q_b > \overline{p} \cdot q_{pb}$ if $q_{pb} > 0$.

Let us name the left hand side of (46) as X:

$$X = \int_{\bar{p}}^{\infty} q_b(\theta, b^W) d\theta$$

Notice that X is always positive. Taking the total derivative of X with respect to $\overline{p}$ we have:

$$\frac{dX}{d\overline{p}} = -q_b + \int_{\overline{p}}^{\infty} q_{bb} d\theta \cdot \frac{db^W}{d\overline{p}} < 0$$

Thus, X will be decreasing in the fare level under the assumptions made regarding the demand function and the sufficient conditions for the optimal frequency levels to be decreasing in the fare ($q_{pb} \leq 0$ or $q_b > \overline{p} \cdot q_{pb}$ if $q_{pb} > 0$).

Let us define the right hand side of (46) as *Y*:

$$Y = (n-1) \cdot \left(\bar{p} \cdot \frac{1}{n \cdot b} \cdot q(\bar{p}, n \cdot b) - c \right)$$

Taking the derivative with respect to the fixed fare and evaluating the expression at the social optimal aggregate frequency level and ignoring the (*n-1*) term, which is irrelevant for the sign of the slope, gives:

$$\frac{dY}{d\bar{p}} = \frac{1}{b^W} \cdot \left(q - \bar{p} \cdot q_b\right) + \frac{1}{b^W} \cdot \left(q_b - \frac{q}{b^W} \right) \cdot \frac{db^W}{d\bar{p}}$$

The term in the first parenthesis will be positive if demand is inelastic for all relevant fare levels. The second parenthesis is the difference between the demand change when frequency increases and the average demand per bus. If $q_b < {}^{q}\!/_{b^W}$ then the second term of the above expression will be positive and $\frac{dY}{d\bar{p}} > 0$.