Economics of Transportation 34 (2023) 100312

Contents lists available at ScienceDirect

Economics of Transportation

journal homepage: www.elsevier.com/locate/ecotra

Subsidised transport services in a fiscal federation: Why local governments may be against decentralised service provision

Daniel Hörcher [a,*], Bruno De Borger [b], Daniel J. Graham [a]

[a] Transport Strategy Centre, Department of Civil and Environmental Engineering, Imperial College London, United Kingdom
[b] Department of Economics, University of Antwerp, Belgium

ARTICLE INFO

JEL classification:
H70
H42
H71
H77
R13
R40
R48

Keywords:
Fiscal federalism
Local public services
Efficient pricing
Tax exporting
Transport subsidies

ABSTRACT

In this paper we consider a fiscal federation and study the effects of decentralised provision of *loss-generating* public services with benefit spillovers to other regions. We use public transport provision across administrative borders as a prototype example. We show in a formal model that local governments might be better off when a higher-level government or a neighbouring region provides these services, and even privatisation to a monopolist can be preferred over decentralisation. Our model reveals that these results are governed by a variant of the tax exporting mechanism that applies to subsidised services, i.e., the possibility that local consumers can exploit spillover benefits without contributing to the subsidy burden of service provision. Public transport provision is one of the large sectors of public policy where decentralisation could provide social benefits, but, as the paper reveals, the need for subsidies generates a genuine conflict of interest between the governments involved.

1. Introduction

The assignment of public service provision in a multi-tier system of government is one of the core subjects of fiscal federalism (Boadway and Shah, 2009; Kitchen et al., 2019). This literature has identified the conditions under which centralisation or decentralisation is the preferred option from the viewpoint of overall social welfare. The leading preference-matching argument supporting decentralisation is the ability of lower-tier governments to better customise the attributes of the service to the preferences of its constituents (Oates, 1972). By contrast, if the service involves substantial spillovers between regions, centralisation might be more efficient, because local governments may neglect spillover effects on other regions. These include classical consumption externalities (such as pollution and congestion) as well as fiscal externalities, for example due to tax exporting (Arnott and Grieson, 1981). Several more recent studies questioned the implicit assumption in the early fiscal federalism literature that central governments are unable to provide differentiated local services (see, for example, Lockwood, 2002; Besley and Coate, 2003; Knight, 2004). Taking into account the political decision making process and allowing such differentiated service provision then clearly weakens the case for centralised decisions (Besley and Coate, 2003).

Importantly, the focus in the above literature has been on whether or not decentralisation is desirable for society as a whole, using aggregate social welfare as the relevant criterion. In this paper we take another look at the assignment of authority to different government levels, emphasising the implications for individual regions. The analysis of local welfare in multi-level governments has received limited attention in the literature so far, probably because it seems obvious that local communities would support decentralisation. In the original framework of Oates (1972) this is a plausible hypothesis, since local communities are the main beneficiaries of the decentralisation of public goods, and their welfare is unaffected by any spillovers to other regions. Moreover, the possibility of tax exporting implies that regions are interested in attracting activities with outside users they can tax. The present paper shows that by the same token, however, regions may avoid producing public services that generate spillover benefits to other regions. This is especially the case if service provision – e.g. due to scale economies – imply financial losses to the region operating the service. Regions may instead prefer that loss making services would be

* Corresponding author.
E-mail address: d.horcher@imperial.ac.uk (D. Hörcher).

https://doi.org/10.1016/j.ecotra.2023.100312
Received 21 October 2021; Received in revised form 19 December 2022; Accepted 20 March 2023
Available online 29 March 2023

operated by other regions or by the federal government; it allows them to benefit from the service offered while avoiding having to finance the subsidies to cover the financial deficit. Public transport may well be the prototype example of this behaviour. For example, a suburban region may prefer that public transport connecting the suburbs and the inner city is organised and financed by the city government. The suburban region benefits from the service offered while the city may carry the largest burden of covering the subsidies. In general, regions may prefer other regions to offer the public transport service even when it implies that the attributes of the service (such as its price and frequency) are sub-optimal from the local perspective.

To the best of our knowledge, this is the first paper to investigate the effects of decentralisation of *loss-making* services with spillovers for individual regional authorities.[1] We consider a federal system with a higher-level 'federal' government and several lower-level 'regional' governments, and we study the provision of a public service that can be provided by each of the different governments. To make this setting more concrete, we study the provision of a public transport service connecting the suburban and urban regions in a large metropolitan area. Both urban and suburban residents commute to a common central business district. Household labour supply requires commuting, and due to agglomeration economies, transport supply has a positive impact on urban productivity and wages. Road and public transport users suffer from congestion and crowding externalities, respectively, and road use imposes further inconveniences on urban residents. There are no congestion tolls; road use is under-priced. The two transport modes are imperfect substitutes so that some of the negative externalities of road use can be neutralised by high-quality and affordable public transport services. Public transport is operated under scale economies, and we consider both the fare and the frequency of service as explicit policy instruments. We then ask the following question: What is for a given region the preferred government to operate the public service under investigation? For example, under what conditions will the urban (suburban) region be better off if the public transport service were operated by the suburban (urban) or federal government?

Within this setting, the model distinguishes four potential regimes of transport provision: (i) centralised, *federal* decision making; (ii) decentralised decision making by the *urban* government; (iii) decentralised decision making by the *suburban* government; (iv) *profit maximising* service provision in monopoly.[2] We study how the choice between these institutional regimes affects the quality and pricing of public transport. More importantly, using these results, we study the preferences of local (urban or suburban) residents in terms of public transport governance regimes.

The analysis shows that, in general, which governance structure is considered most appropriate by different regional communities depends on the extent of benefit spillovers, the magnitude of operational subsidies, and the allocation of the financial burden of the subsidies across regions. On the one hand, the region operating the service has control over pricing and frequency decisions. On the other hand, both the urban and suburban regions benefit from fiscal spillovers when another government provides the service. A free-rider strategy can then be realised in the centralised regime and especially when the neighbouring region becomes responsible for service provision.

Numerical implementation of the model to a generic but not unrealistic example finds that public transport provision leads to a financial deficit in all centralised and decentralised scenarios. Under such circumstances, local residents do not benefit from decentralising the service to their own government, primarily due to the subsidy burden. We find that the urban region achieves higher welfare with suburban control over the service and, more surprisingly, they consider privatisation as superior to their own urban governance regime. The suburban community would prefer federal service provision, because in this case fares are low and service quality is very high, and the subsidy burden can be shared with urbanites from a common tax pool. By contrast, the federal regime is the least preferred alternative for urban residents. These results suggest a potential conflicts of interest between the urban, suburban, and federal governments.

Our analysis is motivated by the diversity we observe in the governance of suburban public transport in cities around the world.[3] For example, in several countries such as Austria, Belgium and Denmark, even purely suburban rail services are operated by national railway companies. In Paris, suburban rail supply is controlled by a transport authority owned by the surrounding Île-de-France region. In several German conurbations, the commuter networks are tendered by regional (state) authorities. The ownership structure of suburban railways may also change over time. For example, the rail network of London's metropolitan area has been part of the national rail franchise system since the 1990s, but Transport for London, owned by the urban government, is now taking control of several lines including Crossrail, a new service linking the Eastern and Western suburbs through the city centre. By contrast, the city of Budapest in Hungary has handed over the suburban rail network it owned to the national railway operator in 2016. These examples suggest that a fiscal federalism approach is highly relevant in the context of public transport provision.

The paper is structured as follows. Section 2 relates the current paper to the literature on fiscal federalism and on the economics of public transport provision. In Section 3 we develop a model and derive analytical pricing and frequency rules for four governance regimes (centralised, urban, suburban, private). In Section 4 we implement these rules in a numerical simulation and investigate the main determinants of local support for decentralisation. Finally, Section 5 concludes.

2. Previous literature

Our study builds on three strands of literature. First, the paper relates to the literature of local public service provision in a fiscal federation. In a system of multi-tiered government, the policy question is twofold: what is the appropriate government to provide the public service, and how will it set the main attributes of the service (user charge, quality level)? The fiscal federalism literature initiated by Oates (1972) has studied these questions, from both normative and political economy perspectives. The normative approach stressed the importance of spillovers between jurisdictions, and it emphasised the role of horizontal (tax exporting) and vertical (overlapping tax bases) fiscal externalities that may emerge under decentralised service provision (see, for example, Dahlby, 1996; Wilson, 1999; Boadway and Shah, 2009). Political economy models showed that the implications for the appropriate level of government to provide a public service may drastically change as soon as the voting behaviour of elected representatives is taken into account; see, among many others, Lockwood (2002), Besley and Coate (2003), Knight (2004), Lörz and Willmann (2005), and infrastructure-oriented analyses in Brueckner and Selod (2006), Glaeser and Ponzetto (2018).[4]

[1] De Borger and Proost (2016) did study the conditions under which regional authorities would deliberately transfer the authority to toll congested regional roads to the federal level.

[2] Profit maximisation serves as a benchmark in the analysis; even where suburban public transport is privately provided (for example in Tokyo and Hong Kong), it is typically heavily regulated by the relevant public authorities.

[3] Due to urban sprawl, millions of suburban commuters rely on interregional transport services on a daily basis. For example, the ratio of the population living within the official city limits relative to the entire metropolitan area is only 36% in Tokyo (13.5 million versus 37.3 million), 42% in Mexico City (8.9 million and 20.9 million), and 61% in London (8.8 million within Greater London and 14.4 million in its metropolitan conurbation).

[4] Voters are typically assumed to be actual or potential users of the public service, but this need not be the case. They may be motivated by social motives such as altruism (Coate, 1995), equity (Gasparini and Pinto, 2006), or social status and reputation (Friedrichsen et al., 2021).

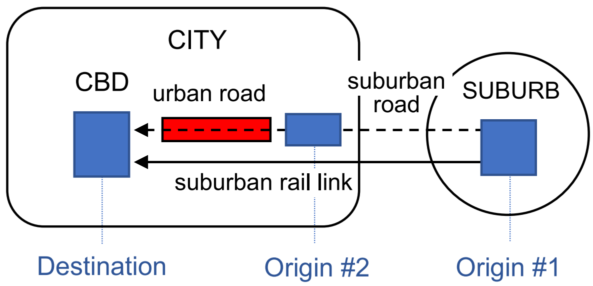

Fig. 1. Schematic network layout.

Second, a number of studies have applied the principles of fiscal federalism and decision-making in multiple-level governments to transport pricing and investment (for an early survey, see De Borger and Proost, 2012). Although models have been developed to study policy competition between ports and between airports, the majority of this literature looks at tax and capacity competition for road transport. Given that the tax base of road pricing is mobile by definition, the regional spillovers that this implies are likely to lead to tax exporting behaviour, either by direct price discrimination against outside users, or by charging higher fees on local roads intensively used by outside users. A number of papers have studied this horizontal fiscal externality on stylised parallel (De Palma and Lindsey, 2000; De Borger et al., 2005) and serial (Levinson, 2000; Van Dender, 2005; De Borger et al., 2007) networks. Moreover, horizontal competition in infrastructure development has been studied by Borck and Wrede (2009) and Vandyck and Proost (2012) in the local context. As decentralised decision making neglects the spillover benefits of local investment, insufficient investment may be undertaken by individual regions. However, if road pricing is in place and the local government is able to attract foreign transport users, excessive investment may result to facilitate tax exporting.[5] Lastly, the political economy of road pricing and public transport provision in a federal system was explored in Xie and Levinson (2009) and De Borger and Proost (2015, 2016).

Third, our model relies on the extensive literature of optimal pricing and frequency decisions in public transport (see Hörcher and Tirachini, 2021, for a comprehensive review). Our model presented in Section 3 extends Hörcher et al. (2020) to multiple regions and multiple levels of government. The model takes into account the costs of waiting time at the station, ensuring scale economies in service provision (Mohring, 1972, 1976; Jansson, 1980). Moreover, it captures both external costs and external benefits of public transport supply. On the one hand, extra passengers per vehicle impose a crowding externality on other passengers (Kraus, 1991; Jara-Díaz and Gschwender, 2003; Tirachini et al., 2013; Hörcher and Graham, 2018). On the other hand, given the complementarity between commuting demand and labour supply (Mayeres and Proost, 1997; Parry and Bento, 2001; Tikoudis et al., 2015), public transport may contribute to urban agglomeration economies (Graham, 2007; Melo et al., 2009; Hörcher et al., 2020). Our analysis further relies on the extensive literature on the substitutability between public transport and under-priced private car use and the potential justification of public transport subsidies (Arnott and Yan, 2000; Ahn, 2009; Parry and Small, 2009; Basso and Silva, 2014). Finally, the model takes into account the distortions that subsidies generate due to the marginal cost of public funds (Parry and Small, 2009; Basso and Silva, 2014; Börjesson et al., 2018).

[5] The literature on transport tax competition between vertically related governments due to overlapping tax bases is much thinner. Examples include Proost and Sen (2006) and Ubbels and Verhoef (2008).

3. The model

The spatial layout of the model is plotted in Fig. 1. We consider a multimodal corridor linking a suburban area with the economic core of a monocentric city. We distinguish two types of commuters and their respective origin–destination pairs for daily travel. The first group consists of N^s *suburban* commuters who reside in the 'suburb' and work in the 'CBD', crossing the boundary between the two municipalities on their way to work. There is a second group of N^u commuters who reside in the same municipality where the employment centre is located. Let us call them *urban* commuters. We assume suburban residents have two options for travelling: a direct rail service to the CBD and private car use, through two serial road links. The key property of this layout is that the inner road link has fixed capacity and is shared between urban and suburban commuters, and therefore they both contribute to the level of congestion inside the city.

We make several assumptions to keep the model simple. First, urban residents have no access to the suburban rail service; in other words, we consider a direct connection between the suburban region and the CBD.[6] Second, we ignore purely local urban bus services, which may be operated in a different institutional setting; as a consequence, urban commuters are all car users. Third, only the urban road section (shared by urban and suburban car users) is congested; the suburban road link has sufficient capacity to accommodate demand without any travel delay. Admittedly, these are strong assumptions, and it is conceptually easy to extend the setup to a shared inter-regional commuter rail service and to allow congestion on the suburban road section. However, this complicates the analytics without offering extra insight into the problem we want to focus on.

3.1. Household behaviour: labour supply and commuting demand

We discuss household behaviour for *suburban* residents. Urban households' preferences and constraints will be specified the same way, except for the fact that they have no access to the inter-regional public commuting service.

The representative suburban consumer's utility is specified as a strictly increasing and concave function of monetary spending on an aggregate consumption good y (its price is normalised at one), effective leisure time l, and the number of commuting trips by rail and car x_r and x_c. Household utility is

$$U = u(y, l) - \phi(x_r, x_c), \quad (1)$$

where the second term $\phi(x_r, x_c)$ captures the disutility of travel, which is independent of consumption and leisure time (this approach follows Parry and Bento, 2001). Consumers face two restrictions. The first one is a monetary budget constraint

$$y = (x_r + x_c)w - x_r p - x_c \tau, \qquad [\kappa] \quad (2)$$

which states that the difference between the total wage income earned (w is the daily wage) and the commuting cost is what is available for consumption; p and τ denote the monetary expense per trip by public transport and private car use, respectively. The Lagrange multiplier of the budget constraint is given in square brackets. The second constraint is a time restriction:

$$l = \bar{L} - (x_r + x_c)L - x_r g_r(f, Q_r) - x_c g_c(Q_c). \qquad [\mu] \quad (3)$$

[6] A series of sensitivity tests are performed in the Appendix and summarised in Section 4.5. Importantly, in Appendix D we allow urban commuters to have access to the suburban rail service as well. We find that this did not affect the main conclusions of the paper, see Appendix D below.

In this expression, $\bar{L}$ is the total time endowment, L is the exogenous daily working hours[7] (e.g. eight hours), $g_r(f, Q_r)$ is the *generalised* travel time by rail, and $g_c(Q_c)$ is the generalised travel time by car. Note that $g_r(f, Q_r)$ and $g_c(Q_c)$ include the travel time equivalent of crowding[8] and congestion-dependent discomfort. In other words, we interpret (3) as an effective leisure time constraint; it not only takes into account the actual time needed for the commuting trip, but also the fact that time spent in a crowded public transport vehicle or in heavily congested car traffic makes consumers more exhausted and cuts into the remaining effective leisure. Train service frequency f and total rail demand Q_r determine the waiting time and crowding experience of passengers; $\partial g_r/\partial f < 0$ and $\partial g_r/\partial Q_r > 0$. Q_c is total car demand and thus a measure of road congestion with $g_c'(Q_c) > 0$. Aggregate traffic Q_c includes both urban and suburban car commuters; see Section 3.2 for more details.

Households maximise (1) subject to (2) and (3), treating g_r and g_c as exogenous. First-order conditions of this problem are

$$\begin{aligned} &u_y - \kappa = 0; \quad u_l - \mu = 0; \\ &\kappa(w - p) - \mu(L + g_r) - \phi_{x_r} = 0; \\ &\kappa(w - \tau) - \mu(L + g_c) - \phi_{x_c} = 0, \end{aligned} \tag{4}$$

where subscripts of u and ϕ denote partial derivatives. Increased commuting (and hence days worked) yields higher monetary income for the household, at the expense of leisure time and the monetary costs and inconvenience of travelling. The solution of (4) leads to individual travel demand functions

$$x_r(p, f, \tau\,;\, w, Q_r, Q_c) \quad \text{and} \quad x_c(p, f, \tau\,;\, w, Q_r, Q_c). \tag{5}$$

The resulting indirect utility is

$$V(p, f, \tau\,;\, w, Q_r, Q_c). \tag{6}$$

In these expressions, p, f and τ are transport supply variables controlled by a federal, regional, or private transport provider. The equilibrium wage and aggregate transport demand levels are further determinants of individual demand and indirect utility through travel times and other travel inconveniences.

As mentioned previously, our modelling framework includes two origin–destination pairs according to the network layout in Fig. 1. The behaviour of urban households is assumed to follow the same rules, except for the fact that their only mode of commuting is the car. Their time and money constraints are equivalent to (2) and (3) after the terms referring to rail commuting are removed. We do allow for other parameters of utility to vary between urban and suburban commuters, so that the preferences of the two groups might differ.

3.2. Commuting, agglomeration, and urban productivity

As mentioned above, the urban road section is congestible, and public transport may suffer from crowding. This subsection explains how agglomeration benefits were introduced into the model. Let superscripts u and s represent the urban and suburban groups of commuters. In the network layout of Fig. 1, urban and suburban drivers share the inner road section, while the rail service is used by suburban residents only:

$$Q_c = N^s x_c^s + N^u x_c^u; \quad Q_r = N^s x_r^s. \tag{7}$$

In these expressions, N^u and N^s refer to the population of the urban and suburban regions, respectively.

Firms in the CBD produce a homogeneous output under constant returns to firm size, using labour as the single production factor. We assume that productivity depends on aggregate labour supply; it has a positive impact on firms' ability to share resources, and it allows exploiting knowledge spillovers and matching jobs with appropriately skilled employees. Firms cannot affect worker productivity internally so that the size of individual firms has no role in the model. We assume perfect competition between profit maximising producers and competition on the labour market. This implies that commuters are paid the value of the (internal) marginal product of the labour they supply.

We capture agglomeration economies by specifying the wage as follows:

$$w = w^0 \cdot (Q_r + Q_c)^{\eta - 1}, \tag{8}$$

where η is the elasticity of output with respect to total labour supply and w^0 is the base value of the marginal product if there are no agglomeration economies ($\eta = 1$). In sum, the wage is endogenous with respect to total labour supply due to the productivity benefits of agglomeration. Therefore, suburban and urban residents are mutually affected by changes in the other group's labour supply.

Urban and suburban residents have different interests in the interregional rail service. Suburban residents are users, so they are interested in low fares and high frequency. Urban residents are not using the service, but they have at least two reasons to support interregional rail. First, it reduces congestion inside the city, reducing their commuting costs. Second, public transport provision is just another way to boost labour supply; given agglomeration economies this raises productivity and the equilibrium wage rate.

3.3. Optimal transport supply

The transport provider's fare and frequency decisions affect the well-being of both urban and suburban commuters, but urban and suburban local governments represent the interests of only a subset of commuters affected. In what follows we consider the objectives of the federal, urban and suburban governments; for the sake of completeness, we also consider a profit-maximising private firm.

Federal governance

The federal government is assumed to maximise aggregate social welfare putting equal weight on the individual utility of urban and suburban users. Its objective is

$$\begin{aligned} \max_{p,f} \Lambda =& N^s V^s(p, f, \tau) + N^u V^u(p, f, \tau) \\ &+ \beta[Q_r \cdot p + Q_c \cdot \tau - z(f)]. \end{aligned} \tag{9}$$

The first two terms in (9) measure consumer surplus by adding up the total indirect utility of the two groups. The final term has the net financial result of transport provision in square brackets. Revenues come from two sources: fare revenues from public transport provision and toll revenues from road usage. We assume that the fare is a decision variable for the federal government. However, the road toll, if any, is assumed to be exogenous, although it does affect the demand for car and public transport use.[9] The last component, $z(f)$ captures the operations cost of providing the rail service; this cost is a function of

[7] Just like Parry and Bento (2001) and others in the literature, we assume that workdays have a fixed length, the representative user commutes by both modes, and aggregate labour supply is proportional to the number of commuting trips, thus creating a direct link between travel demand and labour supply. Hörcher et al. (2020) discuss the validity of, and empirical support for, this assumption in more detail.

[8] Expressing the inconvenience of crowding in terms of the equivalent travel time is a standard modelling practice in the public transport literature (Hörcher and Tirachini, 2021). Empirical evidence on crowding dependent travel time multipliers is reviewed in Wardman and Whelan (2011) and Li and Hensher (2011).

[9] Note that we do not treat the toll as a decision variable in our analysis. One justification for this assumption might be that public transport and road tolls are operated by separate federal agencies.

the frequency offered. The budgetary result is then multiplied by β, the marginal social value of income. This parameter transforms the surplus or deficit into social utility (see e.g. Van Dender, 2003). We assume β is exogenous and known in the optimisation problem.[10] This leads to an objective function in utility terms. In alternative governance regimes we will allow β to vary between various levels of governments to represent differences in their ability to raise taxes in their jurisdictions.

Section A.1 in the Appendix provides first-order conditions for (9) and a series of algebraic manipulations which lead to the following optimal pricing rule:

$$p = MC_r + \left(MC_c - \tau\right)\frac{\frac{\partial x_c^s}{\partial p}}{\frac{\partial x_r^s}{\partial p}} + \left(MC_c - \tau\right)\frac{N^u}{N^s}\frac{\frac{\partial x_c^u}{\partial p}}{\frac{\partial x_r^s}{\partial p}} + \frac{\kappa^s - \beta}{\beta}\frac{x_r^s}{\frac{\partial x_r^s}{\partial x_p}}, \tag{10}$$

where MC_r and MC_c denote the marginal social costs of rail and car commuting trips. The former is given by

$$MC_r = \frac{\mu^s}{\beta}N^s x_r^s \cdot \frac{\partial g_r}{\partial Q_r} - \left(\frac{\kappa^s}{\beta}N^s\left(x_r^s + x_c^s\right) + \frac{\kappa^u}{\beta}N^u x_c^u\right)\cdot w'. \tag{11}$$

In Eq. (11) and through the rest of the paper f' denotes the first derivative of any function f. The marginal social cost of rail commuting includes the social value of the marginal external crowding cost imposed on other rail users. Moreover, it takes account of a positive productivity externality that every commuter enjoys: a marginal increase in labour supply implies a marginal increase in the wage paid. The marginal social cost of car use also has two main components.

$$\begin{aligned} MC_c = &\sum_{i\in(s,u)}\frac{\mu^i}{\beta}N^i x_c^i g_c'(Q_c) \\ &- \left(\frac{\kappa^s}{\beta}N^s\left(x_r^s + x_c^s\right) + \frac{\kappa^u}{\beta}N^u x_c^u\right)\cdot w'. \end{aligned} \tag{12}$$

The first term is the marginal external congestion delay and inconvenience cost imposed on suburban and urban drivers. The second component is the same agglomeration benefit as in MC_r, as the model assumes that mode choice has no impact on commuters' contributions to urban productivity.

Interpretation of the pricing rule in (10) is straightforward. The first term is (11). It captures the (negative) crowding and (positive) productivity externalities of rail use, as explained above. The second and third terms reflect the second-best role that public transport fares may have if road pricing deviates from the net marginal social cost of car trips. The degree to which such distortions in the road market matters for rail pricing depends to the strength of modal substitution. In the absence of modal substitution or with optimal road pricing ($\tau = MC_c$), the second and third terms disappear from the formula. The final term of (10) measures the operator's revenue generating pressure due to the marginal cost of public funds. If there are no distortions elsewhere in the tax system, and therefore $\kappa^s = \beta$, so the government evaluates money the same way as households, then this mark-up would also disappear from the pricing rule. By contrast, if the government assigns much higher value to money than suburban residents, so that $\beta \gg \kappa^s$, then the fraction $(\kappa^s - \beta)/\beta$ approaches -1, and the last term in (10) becomes a pure monopoly mark-up.

The analytical formula for the welfare-maximising service frequency is derived in Section A.2 of Appendix. The frequency rule in Eq. (A.8) follows traditional results in the literature of public transport economics.

Suburban governance

Under suburban service provision, the objective function modifies in three ways compared to (9). First, we assume the suburban government does not consider the welfare of urban residents. Second, the public transport surplus or deficit now enters the suburban government's budget; its marginal value to the local government is denoted β^s. A third modification concerns the treatment of the revenues of road tolls. Although other formulations are possible, we assume that road tax revenues remain with the federal government, but the suburban supplier takes into account that a θ^s share of the welfare effect of this federal revenue will benefit suburban residents. To understand the treatment of toll revenues, we assume that these revenues relax the federal government's budget constraint, and this is perceived as a welfare gain by all residents of the federation because other distortionary federal taxes can thus be reduced. The suburban government does not receive a direct inter-governmental transfer in the model (road tax revenues are not redistributed directly), but it does take the aforementioned welfare gain of its residents into account.[11] In sum, we specify the suburban government's decision problem as follows:

$$\max_{p,f} \Lambda^s = N^s V^s(p,f,\tau) + \beta^s[Q_r \cdot p - z(f)] + \theta^s \beta Q_c \cdot \tau. \tag{13}$$

The first-order condition of (13) with respect to the rail fare yields the following expression, after rearrangement.

$$\begin{aligned} p = &MC_r^s + \left(MC_c^s - \tau\right)\frac{\frac{\partial x_c^s}{\partial p}}{\frac{\partial x_r^s}{\partial p}} \\ &+ \left(MC_c^s - \frac{\theta^s\beta}{\beta^s}\tau\right)\frac{N^u}{N^s}\frac{\frac{\partial x_c^u}{\partial p}}{\frac{\partial x_r^s}{\partial p}} + \frac{\kappa - \beta^s}{\beta^s}\frac{x_r^s}{\frac{\partial x_r^s}{\partial p}}. \end{aligned} \tag{14}$$

The structure of the fare formula is similar to the federal government's optimal pricing rule in (10). However, the difference is that MC_r^s and MC_c^s are the net marginal costs that rail and car travellers impose on the welfare of *suburban* residents only. In particular,

$$MC_r^s = \frac{\mu^s}{\beta^s}N^s x_r^s \cdot \frac{\partial g_r}{\partial Q_r} - \frac{\kappa^s}{\beta^s}N^s(x_r^s + x_c^s)\cdot w' \tag{15}$$

and

$$MC_c^s = \frac{\mu^s}{\beta^s}N^s x_c^s g_c'(Q_c) - \frac{\kappa^s}{\beta^s}N^s(x_r^s + x_c^s)\cdot w' \tag{16}$$

These net marginal cost expressions do not include the external congestion costs and agglomeration benefits borne or enjoyed by urban households. This implies that if the road toll is set to its *social* optimum, it is likely that $MC_c^s \neq \tau$, and therefore the suburban operator's rail fare would deviate even further from the socially optimal p. The sign of this deviation remains ambiguous in the analytical model. The suburban government's optimal frequency rule is discussed in Section A.3 of Appendix.

Urban governance

The urban government takes into account the utility of urban residents only. By analogy to the suburban case, the problem facing the urban government is formulated as follows.

$$\max_{p,f} \Lambda^u = N^u V^u(p,f,\tau) + \beta^u[Q_r \cdot p - z(f)] + \theta^u \beta Q_c \cdot \tau, \tag{17}$$

where β^u is the multiplier of the financial result of service provision in its objective function. The interpretation for the treatment of toll revenues is the same as before as well: road tax revenues relax the federal budget, and urban residents, as members of society, experience a θ^u share of the corresponding welfare gain.

Recall that urban residents are not users of the suburban rail service. The implications of this assumption on the urban government's optimal

[10] Mathematically, the optimality conditions derived from this objective function are formally identical to the optimisation of a Lagrangian function associated with welfare maximisation subject to a budget constraint, $\Lambda = \sum_i N^i V^i - \beta\left[Q_r p + Q_c \tau - z(f) - M\right]$, where β is an endogenous Lagrange multiplier determined by the specific value of an explicit budget limit M. Indeed, in certain policy situations, public transport operators face an explicit budget constraint.

[11] As mentioned, our assumption with respect to the treatment of toll revenues is just one of several other possible scenarios. Note that in the special case of $\beta^s = \beta$, our results are equivalent to the case where federal revenues are redistributed to local governments.

decisions are clear (see Section 4.5 and Appendix D). The marginal impact of rail demand on urban residents' generalised consumption $\partial V^u/\partial x_r^s$, without considering induced demand effects, is limited to the agglomeration benefit of increased labour supply. We denote this marginal *benefit* of rail trips as MB_r^u. Suburban road trips, however, may still impose marginal external congestion costs on the urban household. Thus,

$$\begin{aligned} MB_r^u &= \frac{\kappa^u}{\beta^u} N^u x_c^u \cdot w', \\ MC_c^u &= \frac{\mu^u}{\beta^u} N^u x_c^u \cdot g_c'(Q_c) - \frac{\kappa^u}{\beta^u} N^u x_c^u \cdot w'. \end{aligned} \tag{18}$$

Using similar derivations as before, the optimal fare rule of the urban provider becomes

$$p = -MB_r^u + \left(MC_c^u - \frac{\theta^u \beta}{\beta^u}\tau \right) \left(\frac{\frac{\partial x_c^s}{\partial p}}{\frac{\partial x_r^s}{\partial p}} + \frac{N^u}{N^s} \frac{\frac{\partial x_c^u}{\partial p}}{\frac{\partial x_r^s}{\partial p}} \right) - \frac{x_r^s}{\frac{\partial x_r^s}{\partial x_p}}. \tag{19}$$

In the order of appearance in (19), the optimal fare has the following components. First, the agglomeration benefit generated by the marginal suburban rail commuter for urban workers reduces the urban government's preferred fare. Second, if the road toll is higher (lower) than the marginal external congestion cost imposed on urban drivers, and there is substitution between suburban rail and car usage, then the fare moves downwards (upwards) to correct for the resulting distortion. The third element applies the same correction if rail supply has an induced demand effect on urban car usage as well. This effect may emerge due to the congestion interaction between urban and suburban drivers, and is expected to be weak. Fourth, there is also a monopoly mark-up in (19). As opposed to previous fare formulae in (10) and (14), the mark-up is now unaffected by the cost of public funds: the urban government applies a mark-up similar to the profit maximising monopolist because residents of its own constituency are not actual users of the service. The optimal rail service rule is elaborated in Appendix A.4.

Private service provision

The last governance regime we consider is that of a profit maximising monopolistic firm. The problem it solves is

$$\max_{p,f} \Pi = Q_r \cdot p - z(f). \tag{20}$$

The private operator of course neglects the amount of road tax revenues collected by federal agencies. The first-order condition with respect to the fare yields

$$p = -\frac{x_r^s}{\frac{\partial x_r^s}{\partial p}}. \tag{21}$$

The profit maximising fare is the well-known monopoly price.

3.4. Overview of analytical pricing rules

Table 1 summarises what additive components appear in the optimal pricing rules in Eqs. (10), (14), (19), and (21). We also indicate the expected sign of each component.

We observe that the 'federal' pricing rule includes all marginal external costs and benefits of commuting and a correction for the cost of public funds. The suburban and urban governance scenarios are indeed intermediate solutions between welfare and profit maximisation. We also note that in the absence of modal substitution between suburban car and rail use, the fare is expected to be higher in all regimes, due to the absence of congestion externalities. By contrast, in the presence of perfect substitution, the fare should be higher because agglomeration benefits in one mode are neutralised by fewer commuters using the other, and therefore the net impact of inducing rail trips on agglomeration drops to zero (Hörcher et al., 2020).

Table 1
Optimal fare components under welfare maximising, decentralised and private transport supply. Signs indicate the direction of quantitative impact on optimal fares.

Fare components (marginal values)	Governance regimes			
	Federal	Suburban	Urban	Private
Rail crowding externality	+	+	0	0
Road congestion externality on suburban users	–	–	0	0
Road congestion externality on urban users	–	0	–	0
Agglomeration benefit for suburban residents	–	–	0	0
Agglomeration benefit for urban residents	–	0	–	0
Mark-up due to costly public funds	+	+	0	0
Revenue maximising mark-up	0	0	+	+

4. Numerical illustration

In this section, we are interested in the implications of different governance structures for the local welfare of urban and suburban households. In other words, which governance structure would households in the two regions prefer if they actually had the choice? To learn more about the quantitative properties of centralised, decentralised and private service delivery we move to a numerical version of the model.

4.1. Model specification

To facilitate the calibration of the numerical model and make use of a number of parameters directly available in the literature, we specify $u(y, l) = \kappa y + \mu l$, and introduce $\Gamma(x_r, x_c) = \kappa^{-1}\phi(x_r, x_c)$ as a monetary equivalent of consumption and leisure time independent modal preferences. Thus, the consumer problem in Section 3.1, i.e. maximising (1) subject to (2) and (3), is equivalent to[12]

$$\begin{aligned} \max_{x_r, x_c} \Lambda^h &= (w-p)x_r + (w-\tau)x_c \\ &\quad + \sigma[\bar{L} - (L+g_r)x_r - (L+g_c)x_c] - \Gamma(x_r, x_c). \end{aligned} \tag{22}$$

In this specification $\sigma = \mu/\kappa$ is the marginal value of time.[13] We adopt this formulation because empirical estimates of the monetary value of time are widely available in the transport literature, which simplifies the calibration exercise. The calibration of $\Gamma(x_r, x_c)$ ensures that the demand system complies with base own and cross demand elasticities available in the literature. To provide sufficient flexibility, we specify it as a quadratic polynomial.[14]

The generalised time of rail trips includes the cost of in-vehicle time, waiting time and crowding; it is specified as:

$$g_r(f, Q_r) = \varphi_w 0.5 f^{-1} + t_r^0 \left(1 + \varphi_c \frac{Q_r}{fS}\right), \tag{23}$$

in which φ_w is the unit value of waiting time, and t_r^0 is the in-vehicle travel time assumed to remain constant. S denotes vehicle size, i.e. a measure of the interior capacity of vehicles. Thus, $Q_r/(fS)$ is the average vehicle occupancy rate, and φ_c is the slope of the crowding-dependent multiplier of travel time. The specification of the time cost

[12] Superscripts s denoting suburban households have been suppressed, as in Eqs. (1)–(3).

[13] Expression (22) is essentially the net generalised income of the household, where the monetary, time and inconvenience costs of commuting are subtracted from the sum of the monetary wage and the value of time endowment $\bar{L}$.

[14] The functional form we use is $\Gamma(x_r^i, x_c^i) = \sum_j \alpha_j^i x_j^i + \sum_j 0.5\gamma_j^i (x_j^i)^2 + \sum_j \delta_j^i x_j^i (x_{\bar{j}\neq j}^i)^2, \forall i = (s,u)$ and $j, \bar{j} = (r,c)$. We set parameters to replicate the demand elasticities outlined in Table 2, presented below. Calibration yielded $\alpha_r^s = -59.3$, $\alpha_c^s = -146.1$, $\gamma_r^s = 132.7$, $\gamma_c^s = 369.6$, $\delta_r^s = 99.5$, $\delta_c^s = 402.2$ for suburban commuters and $\alpha_c^u = -11.2$ and $\gamma_c^u = 80.8$ for urbanites.

Table 2
Baseline simulation parameters.

Variable	Value	Unit	Description
Local economy			
N^s	5,000		Number of suburban residents
N^u	5,000		Number of urban residents
L	8	h/day	Exogenous daily working hours
$\bar{L}$	12	h	Total time endowment
σ	10	\$/h	Value of travel time savings
β	1.1	–	Marginal cost of public funds (MCPF), *federal* government
β^u	1.1	–	MCPF, *urban* government
β^s	1.1	–	MCPF, *suburban* government
w	100	\$/day	Daily wage in initial equilibrium
η	1.04		Agglomeration elasticity
Initial transport supply			
f	10	trains/h	Rail frequency
p_r	8	\$/trip	Rail fare of a round-trip
τ	5	\$/trip	Car tax of a round-trip
S		m^2/train	Vehicle size (calibrated for a base crowding density)
$Q_r/(fS)$	3	pass/m^2	Average vehicle occupancy rate (crowding density)
φ_c	0.15		Slope of crowding multiplier
φ_w	1.2		Multiplier for waiting time
t_r^0	0.5	h	In-vehicle rail travel time
t_c^0	0.25	h	Uncongested travel time by car
ρ	0.6		Self-financing ratio in initial equilibrium
ϵ	0.8		Operational cost elasticity w.r.t. output
Initial demand			
x_r^s	0.25	trip/day	Suburban rail demand
x_c^s	0.25	trip/day	Suburban car demand
x_c^u	0.5	trip/day	Urban car demand
ϵ_{r,p_r}	−0.5		Own-price elasticity of suburban rail demand
$\epsilon_{r,f}$	−0.3		Headway elasticity of suburban rail demand
ϵ_{c,p_r}	0.2		Cross-price elasticity of suburban car use
$\epsilon_{c,f}$	0.15		Cross-headway elasticity of suburban car use
$\epsilon_c{}^u$	−0.3		Travel time elasticity of urban car use

of car use follows the well-known Bureau of Public Roads congestion technology:

$$g_c(Q_c) = t_c^0 \left[1 + 0.15\left(Q_c/K\right)^4\right]. \tag{24}$$

In this formula, t_c^0 is the uncongested travel time, and K is the theoretical capacity of the road.

We derive the utility maximising commuting flows by optimising (22) with respect to x_r and x_c, and updating g_r, g_c and the wage rate w iteratively. The utility of *urban* households is equivalent to (22), but removes all rail commuting related components.

4.2. Policy evaluation

With the functional specifications above, the objectives defined in (9), (13), (17) and (20) for the federal, suburban, urban and private supply regimes are evaluated and optimised with a standard numerical optimisation algorithm. This gives the fare and the frequency under each of the four governance structures, together with the associated welfare for the authority in charge of public transport provision.

Our objective is to evaluate the local welfare of both regions under all four governance regimes. We replace θ_s and θ_u, that is the share of region S and U in federal finances, with their share in total population, N^s/N and N^u/N. As mentioned before, the tax on car use is collected by the federal government in all regimes,[15] and this tax revenue remains with the federal government. However, urban and suburban residents are shareholders in the federal government as well, and therefore federal tax revenues and the corresponding tax revenue premia do affect welfare of local residents. We assume this welfare gain is also split between the two groups (urban, suburban) of users according to their population ratios. Furthermore, we keep the marginal cost of public funds differentiated between governments.[16]

According to the assumptions above, the representative *suburban* households' welfare function can be written as

$$\begin{aligned} W_{s:f} &= N^s V^s(f,p) + \frac{N^s}{N}\beta\left[Q_r\,p + Q_c\,\tau - z(f)\right]; \\ W_{s:s} &= N^s V^s(f,p) + \beta^s\left[Q_r\,p - z(f)\right] + \frac{N^s}{N}\beta\,Q_c\,\tau; \\ W_{s:u} &= W_{s:\pi} = N^s V^s(f,p) + \frac{N^s}{N}\beta\,Q_c\,\tau. \end{aligned} \tag{25}$$

Note that $W_{s:s}$ is identical to (13), the objective function of the suburban government's optimisation problem. Similarly, *urban* households realise the following supplier-dependent aggregate surplus:

$$\begin{aligned} W_{u:f} &= N^u V^u(f,p) + \frac{N^u}{N}\beta\left[Q_r\,p + Q_c\,\tau - z(f)\right]; \\ W_{u:s} &= W_{u:\pi} = N^u V^u(f,p) + \frac{N^u}{N}\beta\,Q_c\,\tau; \\ W_{u:u} &= N^u V^u(f,p) + \beta^u\left[Q_r\,p - z(f)\right] + \frac{N^u}{N}\beta\,Q_c\,\tau. \end{aligned} \tag{26}$$

[15] In Section 4.5 we discuss the sensitivity of simulation outcomes by setting $\tau = 0$ to exclude road tax revenues from the analysis entirely.

[16] The differentiation of β, β_s and β_u would not be appropriate if (i) frictionless monetary transfers were possible between governments, so that the cost of public funds could be equated instantaneously, or (ii) various levels of government had only one common budget. These assumptions do not hold in most countries, and therefore we believe the differentiation of β's does improve the value of the modelling exercise.

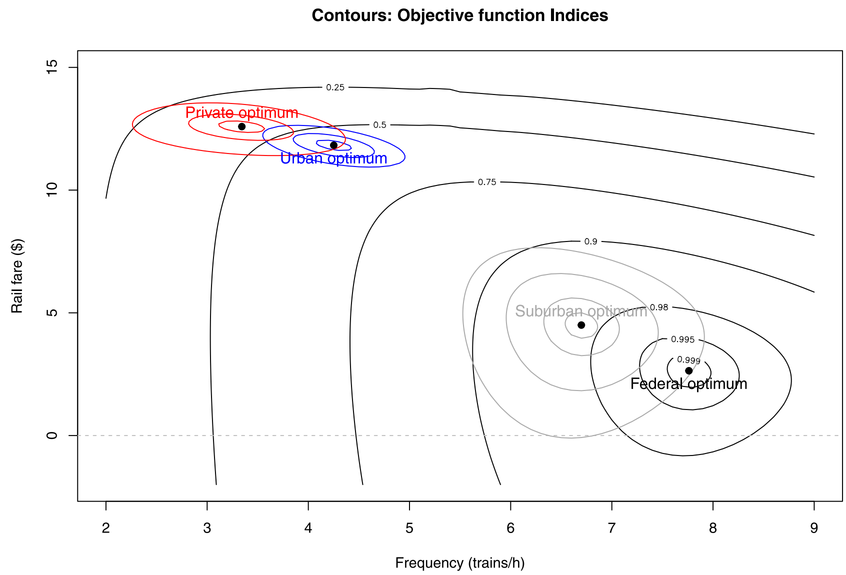

Fig. 2. Optimal supply (rail fare and frequency) under centralised, decentralised and private supply regimes.

Aggregate social welfare when government i is in charge is defined by

$$W_i = W_{s:i} + W_{u:i} \quad \forall i \in \{f, s, u, \pi\}. \tag{27}$$

The financial outcome of public transport provision remains (20).

4.3. Model calibration

The model's exogenous parameters are enlisted in Table 2. Note that vehicle size S is calibrated such that rail crowding in the initial equilibrium is 3 passengers per square metre to represent a peak-hour commuting service. Road capacity K is calibrated to achieve that travel time on the urban road section is twice the uncongested travel time, $g_c(Q_c) = 2 \cdot t_c^0$, where g_c follows Eq. (24). The operational cost function is specified as $z(f) = z^0 f^{\epsilon}$. Elasticity ϵ is set to 0.8 to capture density economies in operational costs, and we calibrate z^0 by prescribing that the cost recovery ratio $\rho = Q_r p / z(f) = 60\%$, which is a typical value in the European public transport context.[17] We specify the value of w^0 in the wage equation of (8) by making sure that $w = \$100$ in the initial equilibrium with agglomeration elasticity $\eta = 1.04$. This value of η is an approximation of the median estimate in the body of literature reviewed by Graham and Gibbons (2019), which is also very close to the estimate recommended by the UK Department for Transport's Transport Analysis Guidance for policy appraisal. The selected value of $\beta = 1.1$, the parameter capturing the marginal cost of public funds, is in the middle or the distribution of international estimates reviewed by Basso and Silva (2014). As the range of empirical evidence is wide, we perform a sensitivity analysis of β, β^u and β^s in Appendix C.

The calibration of the quadratic function $\Gamma(x_r, x_c)$ requires a numerical optimisation exercise. The objective of this optimisation is to minimise the deviation of demand levels and demand elasticities from their initial values provided in Table 2. To ensure that demand sensitivity remains reasonable on a wide range of the supply variables, we simulate 50% perturbations relative to the initial levels $p = \$8$ and $f = 10$ trains/h in the calibration exercise.

[17] More specifically, this is a long-run average of the self-financing ratios of European members within the Community of Metros, a benchmarking group of metro operators facilitated by Imperial College London.

4.4. Results

Let us begin the discussion of the numerical results with the pattern of optimal fare and frequency values under the four supply regimes. The results are graphically depicted in Fig. 2. This graph shows the optimal supply levels identified by the coloured labels. The corresponding coloured contour lines show the rate of reduction in each government's objective function as we deviate from their respective optima.

The relative position of social welfare and profit maximising supply is in line with the literature: the private monopolist sets significantly higher fares and lower frequencies compared to the first-best (federal) optimum. This is primarily due to the monopoly mark-up that the private provider applies. The welfare maximising fare level is influenced by the off-setting effects of (negative) crowding and positive (agglomeration) externalities (Hörcher et al., 2020), which keeps the first-best fare relatively low, despite the presence of substantial crowding (see Fig. 3).

We observe that the decentralised governments' optimal fare and frequency are in between the centralised and privatised ones. This can be explained by the structure of the fare and frequency formulae derived in the theoretical part of the paper. In particular, the *suburban* government does not consider the road congestion externality imposed on urban drivers and the agglomeration benefits of urban residents, while the federal government does (see a summary of theoretical results in Table 1). These two externalities are also crucial when we compare the *urban* government with the private transport provider. In both regimes a revenue maximising mark-up appears in the fare formulae (19) and (21), but the urban government additionally considers road congestion and agglomeration externalities; its optimal rail fare is therefore lower than the profit maximising one.

Fig. 3 zooms in on the implications of different governance structures for crowding in public transport. The contours on the figure correspond to combinations of fare and frequency that imply the same level of crowding, measured by the number of passengers per square metre. The closer the contour is to the horizontal axis, the higher the crowding level. The simulation suggests that despite the significant difference of fare-frequency combinations in the four regimes, equilibrium occupancy rates are fairly similar level, near 4 passengers per square metre. There is slightly more crowding at the federal and suburban optima than under urban governance and profit maximising

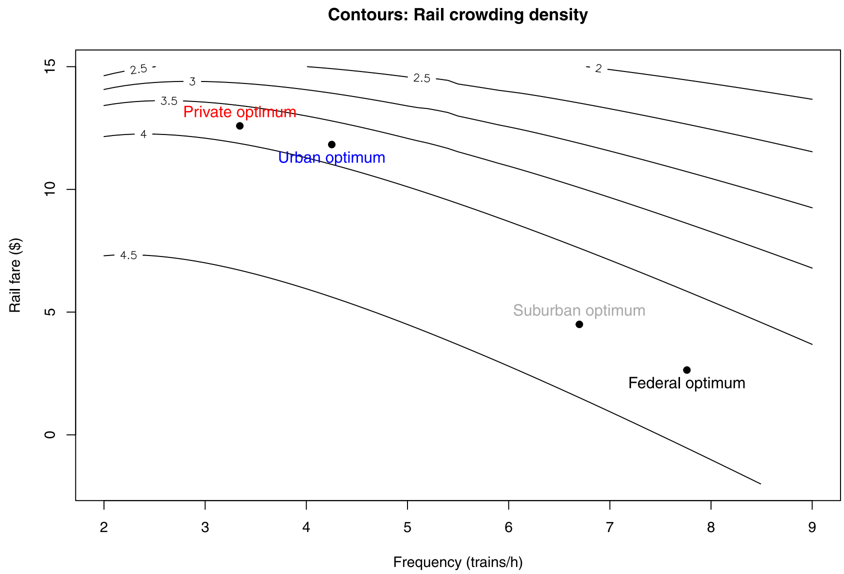

Fig. 3. The pattern of crowding density (in passengers per square metre) and the location of rail supply in four ownership regimes.

operations. This happens because the private and urban suppliers are more interested in keeping rail users' willingness to pay high in order to increase service profitability.

The key numerical results are summarised in Table 3. The first two columns give the numerical values of optimal fares and frequency that were already graphically provided in Fig. 2 above. Using the optimal fare, frequency and equilibrium rail demand levels, in column 3 we compute the self-financing ratio of interregional public transport provision. Under the current set of parameters, the federal regime relies almost entirely on public subsidies, as fare revenues cover only 27.1% of the operational costs. Suburban and especially urban provision achieve higher cost recovery ratios; these authorities have lower costs because they offer lower frequencies, and they charge higher prices than the federal level. The urban government is in fact close to the break even point with a self-financing ratio of 95.1%. Unsurprisingly, the best financial performance is achieved under a profit maximising monopoly.

The welfare results are provided in terms of a welfare index in Table 3:

$$\omega_i = \frac{W_i(f,p) - W_i(0,\infty)}{W_i(f_i^*,p_i^*) - W_i(0,\infty)}, \tag{28}$$

where $W_i(0,\infty)$ is the reference welfare level in the absence of the public transport service ($f = 0$ and $p = \infty$, so that $Q_r = 0$), and $W_i(f_i^*, p_i^*)$ is the maximum welfare obtained when region i is operating the service. Let us first consider the aggregate (or federal) welfare as an example. By definition, the model's 'federal' scenario achieves the highest level of social welfare; the aggregate welfare index we defined before equals one. If the interregional public transport service were provided under the suburban and urban regimes, aggregate welfare would become 97.2% and 57.8% of this maximum efficiency (remember that zero is associated with social welfare in the absence of a suburban public transport service). These findings are not surprising, given that the users of the service are actually suburban residents. With a private monopolist, the aggregate efficiency index drops to 43.1%.

The fiscal federalism literature suggests that decentralisation is often preferred by local residents as it allows their local governments to control the service and capture tax revenues from neighbouring regions. Interestingly, in the present model this is not the case. The column reporting 'suburban welfare' under the four governance regimes in

Table 3
Policy evaluation with aggregate as well as regional welfare indices.

Scenario	Frequency	Fare	Cost recovery	Welfare indices (ω)		
	f (trains/h)	p ($)	$Q_r p/z(f)$	Aggregate	Suburban	Urban
Federal	7.9	2.64	0.271	1	2.160	−1.101
Suburban	6.7	4.50	0.449	0.972	1	2.161
Urban	4.3	11.83	0.951	0.578	0.691	1
Private	3.3	12.59	0.966	0.431	0.470	1.037

Table 3 shows that suburban commuters are better off under centralised service provision than they would be if their own local government organised the public transport connection with the urban region. Under federal public transport provision, the welfare index of suburban residents is more than twice as large as under local suburban governance. When the service were operated by the urban provider, suburban welfare drops to 69.1% of the same basis.

Similarly, urban residents are better off if public transport is not organised by their own government. Their most preferred governance regime is decentralisation, but under the neighbouring suburban region's control. Urban residents gain a lot with a suburban government taking care of the public transport service compared to self-organising the service: the resulting welfare gain is more than twice as high as what the urban government could itself achieve. Remarkably, federal public transport provision is by far the worst option for urban residents. The welfare index is actually negative in this case: urban residents feel worse off under federal service provision than in the complete absence of the public transport service connecting the suburban region with the urban employment area. Finally, another unanticipated result is that privatisation is an attractive alternative for the urban community, as this leads to a 3.7% net welfare gain compared to the 'urban' regime.

To learn more about the main driving forces behind these results, we split the welfare measures defined in (25) and (26) into three categories and compute the impact of public transport provision on each of them. The three categories are:

(i) consumer surplus of public transport provision, i.e. the change in household utility relative to the reference equilibrium in the absence of public transport,

Table 4
Elements of local welfare effects of public transport provision as perceived by suburban (top) and urban (bottom) residents, in each supply regime. Welfare effects are relative to the scenario with no public transport.

Scenario	Consumer Surplus	PT subsidies	Road tax revenue	Total
Suburban residents				
Federal	16.7	−5.5	−1.5	9.8
Suburban	13.1	−7.3	−1.3	4.5
Urban	3.9	0.0	−0.8	3.1
Private	2.8	0.0	−0.6	2.1
Urban residents				
Federal	5.2	−5.5	−1.5	−1.7
Suburban	4.6	0.0	−1.3	3.3
Urban	2.8	−0.4	−0.8	1.5
Private	2.2	0.0	−0.6	1.6

All quantities in $1,000.

(ii) the welfare effect of the passenger group's contribution to public transport subsidies,
(iii) the welfare effect of road tax revenues lost due to the availability of public transport.[18]

In Table 4 consumer surplus is consistent with *a priori* expectations, given the optimal fares and frequencies: federal supply leads to the highest consumer surplus in both groups, and it gradually decreases as we move towards higher fares and lower frequencies in other regimes. The loss in road tax revenues is proportional with the quality of public transport, but its magnitude is significantly lower. Importantly, note that the contribution to public transport subsidies causes huge variations in local welfare results.

Suburban residents enjoy outstanding welfare effects under federal provision because in that regime fares are low and frequencies high, implying high consumer utility; moreover, they share financing the deficit with urban residents. This makes the federal scenario much more attractive for suburbanites than organising the service at the suburban level; in that case they are solely responsible for the deficit incurred. By contrast, the *urban* government would establish an almost self-financing public transport service on its own. In the federal regime they have to contribute to funding a huge subsidy that actually exceeds the consumer surplus they realise from the availability of suburban public transport. In fact, suburban provision of the inter-regional rail link is the most preferred option from the viewpoint of urban residents. In that case, they receive almost the same consumer surplus as in the federal regime, with no further financial burden. Along the same lines, we see that the absence of subsidies is what makes the private regime also relatively attractive for urban commuters.

Our results illustrate an interesting phenomenon: local interest groups (urban and suburban residents) have an incentive to delegate the duty of inter-regional transport provision to federal control or to the neighbouring region's jurisdiction. For example, urban residents want to take advantage from the benefit spillovers of a loss generating service that the neighbouring suburban region subsidises. Similarly, the suburbans would prefer that public transport is operated by the federal level, because in that case the urban population co-subsidises the public transport deficit. This type of behaviour is in a sense the opposite of *tax exporting* behaviour. Tax exporting happens when a competing decentralised government intends to control service provision and realise high revenues paid by the residents of the neighbouring region. What we identified in this paper is a new variant of tax exporting behaviour: regions want to shift the subsidy burden of a loss-generating public transport service to the federal level or to another region. They prefer not to operate the service themselves to avoid a heavy subsidy burden, but they do benefit from its availability and the resulting spillovers.

4.5. Sensitivity analyses

The simulation exercise presented above reveals a number of important properties of decentralised inter-regional public transport provision. We perform a series of sensitivity analyses to check the robustness of our findings with respect to key assumptions and model parameters. Specific details and quantitative results of these sensitivity analyses are provided in Appendix. Here we limit the discussion to a summary of the major results.

In the first exercise in Appendix B we vary the level of exogenous road tax τ. This determines federal tax revenues as well as the potential inefficiencies in road commuting when τ deviates from the marginal social cost of road use. We first consider unpriced road use. In that case, the optimal rail fare drops in all four regimes, in line with the standard second-best argument (Arnott and Yan, 2000). As the urban and federal governments do take the congestion costs of urban commuters into account, they also increase the frequency of rail services. Second, we show that the road tax level is a key determinant of the urban and suburban local governments' behaviour. Low τ implies a heavily subsidised public transport service. We find that suburban residents feel better off under both urban and federal regimes. Urban residents prefer both suburban and private provision. With high road taxes, this motivation disappears due to improved cost recovery under optimal rail pricing.

The second exercise (Appendix C) deals with the possibility that urban road users might generate additional external costs for city residents, beside the congestion externality we already include in the baseline model. Environmental, accident and parking externalities are often mentioned as other adverse effects of urban car use. An extension of the numerical model to account for other external costs does not affect the qualitative conclusions of the paper. Of course, the urban, suburban and federal governments' optimal policies diverge further from the profit maximising fare and frequency; fares decline and frequencies increase to capture higher external costs.

Next, in Appendix D, we relax the assumption that urban commuters have no access to the suburban rail service. This turns the benefit of the urban users of the suburban service into a relevant determinant of supply in the *urban* and *federal* regimes. The consequence of this extension is that suburban commuters may be not necessarily refuse the *urban* regime, provided that the share of the urban population, $N^u/(N^u+N^s)$, is sufficiently large. The opposite finding holds for the preference of urban residents: they no longer prefer the *suburban* regime when $N^u/(N^u+N^s)$ is high, because the suburban government would provide insufficient capacity and high fares for the service mostly used by urbanites. Nevertheless, we find that centralised (federal) supply remains the most preferred regime of at least one of the two interest groups in the entire range of the population distribution.

Finally, the fourth numerical test in Appendix E is concerned with model sensitivity with respect to β, β^u and β^s, i.e. the value of the marginal cost of public funds for the three governments. As it is difficult to determine a priori what the relative magnitude of these parameters should be, we show in a randomised experiment that the frequency and fare optima of the private and urban regimes are not sensitive to the cost of public funds. The suburban and federal regimes react more to changes in β's, primarily by increasing the rail fare when the cost of public funds increases. If the suburban budget is more relaxed than the federal one ($\beta^s < \beta$), then the rail fare can be lower under suburban than under federal control. However, this does not affect the ranking of centralised and decentralised regimes from the viewpoint of local

[18] Note that road tax revenues have a *positive* impact on welfare in (25) and (26), but the introduction of public transport actually *reduces* this tax revenue, and therefore the associated welfare effects in Table 4 are *negative*.

residents' welfare. Thus, we conclude that the paper's main findings are robust with respect to the cost of public funds at various governments.

5. Conclusions

What is for a given region of a fiscal federation the preferred government to operate specific public services?—This is the main research question addressed by this paper. Most of the previous literature on fiscal federalism agreed that preference matching and the possibility of tax exporting makes decentralisation very attractive for regions. We address this question in the context of public transport services connecting multiple political jurisdictions, thus making more than one governments reasonable candidates for service provision. We show that whether residents of the region prefer the service to be operated by their own local government depends on (i) whether or not the service is profitable, (ii) the sign and magnitude of consumption externalities, and (iii) the sharing rules of profits or losses when the outside government is in charge. Most importantly, when optimal pricing leads to a financial deficit, decentralisation is only preferred by a region if an outside government covers a significant part of the deficit. That is, regions may not be in favour of taking the loss-generating service under their control if another government may be ready to fund the loss whilst keeping service quality acceptable. This creates a potential incentive for local governments to act *against* decentralisation. This result is the opposite of findings in previous papers on tax exporting.

In the simulations presented, suburban commuters prefer federal service provision the most, while urban residents would even consider privatisation as superior to their own urban governance regime. These results are governed by the financial loss of a public service characterised by scale economies and the assumed deficit sharing rules. Under different deficit sharing assumptions, some of our qualitative results may change. For example, if a political pact led to a fiscal arrangement in which the suburban region had to cover the entire deficit of federal service provision, then they would clearly prefer decentralisation to the suburban government. Nevertheless, a conflict of interest between the urban, suburban and federal governments is likely to prevail under a wide range of assumptions and parameters, as a series of sensitivity analyses illustrated in this paper.

Our public transport application is based on several context-specific assumptions, some of which could be relaxed or modified to adapt the model to different circumstances. For example, we assume that the tax on individual car use is exogenous. Making the road tax endogenous and jointly optimising policy variables for public and private transport would reflect an alternative policy scenario. This would introduce vertical competition of governments where the multimodal commuting market is the overlapping tax base. Also, the simplified spatial layout of the present paper could be replaced with the standard monocentric city framework (Basso et al., 2021, show the impact of public transport on urban form in this setup), in which an exogenous administrative border between urban and suburban areas is expected to cause discontinuities in commuting costs, real estate prices as well as the structural density of land use. Finally, this paper has taken a normative approach without modelling the institutions and political processes that lead to decisions in the federal and lower-tier governments. A political economy approach could shed light on the impact of the conflict of interest between horizontally related local communities and the way how their interests are represented in central decision making.

CRediT authorship contribution statement

Daniel Hörcher: Conceptualisation, Methodology, Investigation, Methodology, Software, Visualisation, Writing – original draft, Writing – review & editing. **Bruno De Borger:** Conceptualisation, Investigation, Methodology, Writing – original draft, Writing review & editing. **Daniel J. Graham:** Conceptualisation, Funding acquisition, Supervision.

Appendix A. Optimal supply rules

A.1. Pricing under federal service provision

The first order condition of the welfare maximising problem described by the Lagrangian in (9) is

$$\frac{\partial \Lambda}{\partial p} = N^s \frac{\partial V^s}{\partial p} + N^u \frac{\partial V^u}{\partial p} + \beta \cdot \left(Q_r + \frac{\partial Q_r}{\partial p} \cdot p + \frac{\partial Q_c}{\partial p} \cdot \tau \right) = 0, \tag{A.1}$$

The fare affects the indirect utility of both suburban and urban residents, and it affects the government's revenues. Using Eqs. (1) to (4) we show:

$$\begin{aligned} \frac{\partial V^s}{\partial p} &= -\kappa^s x_r^s + \kappa^s (x_r^s + x_c^s) \frac{\partial w}{\partial p} - \mu^s x_r^s \frac{\partial g_r}{\partial Q_r} \frac{\partial Q_r}{\partial p} - \mu^s x_c^s g_c'(Q_c) \frac{\partial Q_c}{\partial p} \\ \frac{\partial V^u}{\partial p} &= \kappa^u x_c^u \frac{\partial w}{\partial p} - \mu^u x_c^u g_c'(Q_c) \frac{\partial Q_c}{\partial p}. \end{aligned} \tag{A.2}$$

The impact of the fare on the endogenous wage and on overall road congestion depends on own and cross-price effects of travel demand. We have:

$$\begin{aligned} \frac{\partial w}{\partial p} &= \left(N^s \frac{\partial x_c^s}{\partial p} + N^s \frac{\partial x_c^s}{\partial p} + N^u \frac{\partial x_c^u}{\partial p} \right) \cdot w' \\ \frac{\partial Q_c}{\partial p} &= N^s \frac{\partial x_c^s}{\partial p} + N^u \frac{\partial x_c^u}{\partial p} \end{aligned} \tag{A.3}$$

Let us now interpret Eqs. (A.2) in more detail. A marginal fare increase affects the representative ***suburban*** user's indirect utility in four ways. First, it increases her monetary payments after each rail trip (causing disutility at a marginal rate κ^s). Second, pricing affects the individual's wage through induced/discouraged commuting demand. Third, induced rail demand generates crowding externalities through the demand-dependent generalised travel time $g_r(f, Q_r)$. Fourth, a mode shift towards car use induces road congestion externalities via generalised time loss $g_c'(Q_c)$. This loss is evaluated at μ^s. By contrast, ***urban*** residents' utility is unaffected by crowding and rail fare payments; see the second row of (A.2). They perceive only the productivity and road congestion externalities; their marginal money, time and travel inconvenience valuations, κ^u, μ^u and $\phi^u_{x_c}$, might also differ from those of the representative suburban user.

Substituting (A.2) and (A.3) into first-order condition (A.1) and rearranging, we obtain the optimal pricing rule reported in Eq. (10) in which the net marginal cost expressions are in (11) for the rail mode and in (12) for car use.

A.2. Optimal frequency under federal service provision

The derivation of the aggregate welfare maximising (i.e. federal) service frequency requires similar analytical steps. The first-order condition of the maximisation of our welfare function (9) with respect to f is

$$\frac{\partial \Lambda}{\partial f} = N^s \frac{\partial V^s}{\partial f} + N^u \frac{\partial V^u}{\partial f} + \beta \cdot \left[\frac{\partial Q_r}{\partial p} \cdot p + \frac{\partial Q_c}{\partial p} \cdot \tau - z'(f) \right] = 0. \tag{A.4}$$

The effect of frequency on the utility of suburban and urban consumers can be expressed as, respectively,

$$\begin{aligned} \frac{\partial V^s}{\partial f} = &- \mu^s x_r^s \left(\frac{\partial g_r(f, Q_r)}{\partial f} + \frac{\partial g_r(f, Q_r)}{\partial Q_r} \frac{\partial Q_r}{\partial f} \right) \\ &+ \kappa^s (x_r^s + x_c^s) \frac{\partial w}{\partial f} - x_c^s \mu^s g_c'(Q_c) \frac{\partial Q_c}{\partial f} \end{aligned} \tag{A.5}$$

and

$$\frac{\partial V^u}{\partial f} = \kappa^u x_c^u \frac{\partial w}{\partial f} - x_c^u \mu^u g_c'(Q_c) \frac{\partial Q_c}{\partial f}. \tag{A.6}$$

Expression (A.5) suggests that suburban residents are affected by f in three ways:

- frequency reduces rail travel time and crowding inconvenience by $\partial g_r(f, Q_r)/\partial f$, but

- it has a feedback effect as well on crowding due to induced demand, i.e. $\frac{\partial g_r}{\partial Q_r}\frac{\partial Q_r}{\partial f}$;
- it may affect the average wage via induced demand;
- it may affect car congestion if the marginal frequency adjustment generates mode shift via $\partial Q_c/\partial f$.

After substituting (A.5) and (A.6) back into (A.4), we reach the following frequency rule.

$$z'(f) = -\frac{\mu^s}{\beta}N^s x_r^s \frac{dg_r(f,Q_r)}{df} - \left(MC_r - p\right)\frac{\partial Q_r}{\partial f} - \left(MC_c - \tau\right)\frac{\partial Q_c}{\partial f} \quad \text{(A.7)}$$

This suggests that frequency should be increased so long as its marginal operational cost is less than the sum of marginal waiting time and crowding benefits, plus the additional welfare effects of induced demand in rail and commuting markets. Induced demand in the two transport markets causes distortions if their respective prices (p for suburban rail and τ for urban and suburban car commuting) deviate from the relevant marginal costs. Frequency has a second-best role if road use is underpriced.

The frequency rule in (A.7) can be transformed into a more familiar form by introducing explicit expressions for the rail travel time function $g_r(f,Q_r)$. Let us split g_r into a frequency-dependent time component, i.e. waiting time t_w, and a crowding dependent element $\Phi(f,Q_r)$, so that $g_r(f,Q_r) = t_w(f) + \Phi(f,Q_r)$. If passengers arrive to their station of origin randomly while train headways are fixed, then the average waiting time is half of the headway, such that $t_w(f) = 0.5f^{-1}$. Consequently, $t'_w(f) = -0.5f^{-2}$. With this standard specification, the frequency rule in (A.7) can be rearranged to

$$f = \sqrt{\frac{\frac{\mu^s}{\beta}0.5\,Q_r}{z'(f) + \frac{\mu^s}{\beta}N^s x_r^s\frac{d\Phi}{df} + \left(MC_r - p\right)\frac{\partial Q_r}{\partial f} + \left(MC_c - \tau\right)\frac{\partial Q_c}{\partial f}}}. \quad \text{(A.8)}$$

This resembles the well-known square root formula of optimal frequency derived many times since Mohring (1972): frequency is proportional to the square root of rail demand Q_r. The numerator of the formula measures the social value of the waiting time benefit of frequency adjustment. In the denominator, we have the marginal operational cost, the marginal impact of f on crowding inconvenience, and the value of further distortions in the rail and car commuting markets that matter if (i) p and τ are sub-optimal, and (ii) frequency has a non-zero impact on rail and car demand.

A.3. Optimal frequency under the suburban government

We derive the suburban government's frequency rule the same way as in the previous case, taking first-order conditions of (13) with respect to f, we find

$$z'(f) = -\frac{\mu^s}{\beta^s}N^s x_r^s \frac{dg_r(f,Q_r)}{df} - \left(MC_r^s - p\right)\frac{\partial Q_r}{\partial f} - \left(MC_c^s - \frac{\theta^s\beta}{\beta^s}\tau\right)\frac{\partial Q_c}{\partial f}. \quad \text{(A.9)}$$

The frequency rule does not differ very much from (A.7). The rail service is used by suburban residents only, and therefore marginal waiting time and crowding benefits are considered by the suburban operator the same way as the centralised government. The only difference is that the marginal welfare effects of induced demand are based on MC_r^s and MC_c^s instead of the full marginal social costs considered by the federal level. As the marginal waiting time saving is still $0.5f^{-2}$, (A.9) can be reformulated as a Mohring-type square root expression:

$$f = \sqrt{\frac{\frac{\mu^s}{\beta^s}0.5\,Q_r}{z'(f) + \frac{\phi_{k_r}^s}{\beta^s}N^s x_r^s\frac{d\Phi}{df} + \left(MC_r^s - p\right)\frac{\partial Q_r}{\partial f} + \left(MC_c^s - \frac{\theta^s\beta}{\beta^s}\tau\right)\frac{\partial Q_c}{\partial f}}} \quad \text{(A.10)}$$

The interpretation of (A.10) is equivalent to (A.8).

A.4. Optimal frequency for the urban government

The first-order condition with respect to the frequency yields the following rule:

$$z'(f) = -\left(-MB_r^u - p\right)\frac{\partial Q_r}{\partial f} - \left(MC_c^u - \frac{\theta^u\beta}{\beta^u}\tau\right)\frac{\partial Q_c}{\partial f}. \quad \text{(A.11)}$$

We do recognise several components of (A.11) from comparing with (A.7) and (A.9). However, note that the marginal waiting time and crowding benefits of frequency adjustment are missing from the urban government's frequency decision as urban commuters do not enjoy these benefits. Thus, the urban service provider increases frequency as long as its marginal financial and welfare effects (i.e. agglomeration benefits and congestion mitigation) outweigh the marginal operational cost. In the absence of direct waiting time benefits, f does not appear explicitly in (A.11), and we can no longer derive the usual square root formula.

A.5. Optimal frequency for the private monopolist

The first-order condition of (20) yields a simple frequency rule.

$$z'(f) = \frac{\partial Q_r}{\partial f}\cdot p \quad \text{(A.12)}$$

in which on the right hand side we have the revenue effects of induced demand. That is, the operator raises frequency as long as the marginal fare revenue is greater than the marginal operational cost. The literature suggests that the magnitude of induced demand, i.e. $\partial x_r^s/\partial f$, also depends on user externalities such as crowding and congestion delay (see Small and Verhoef, 2007, Section 6.1), but these effects are not explicitly visible in the present pricing and capacity formulae.

Appendix B. Sensitivity test: Exogenous road taxes

In the numerical analysis of the paper we assume that the tax imposed on urban car users is exogenously fixed at τ = \$6 for a roundtrip. Of course, the theory of Section 3 showed that for several regimes the optimal rail fair is affected by the gap between τ and the marginal social cost of car commuting. This implies that the actual level of an exogenous road/fuel tax might have a substantial impact on (i) the optimal rail supply, (ii) road tax revenues, and (iii) the benefits of rail service provision.

In Fig. B.1 we first re-run the numerical model simulating the absence of road taxes, i.e. τ = \$0. The dashed contour lines are the ones corresponding to the baseline plotted in Fig. 2, while the solid contours refer to the new optima when the tax is zero. The general finding is that the optimal rail fare decreases consistently in all four supply regimes. This is a departure from the baseline optimum in the expected direction: as the gap between τ and the positive marginal external cost of driving increases, the second-best rail fare declines to compensate for part of the deadweight loss of underpriced road use.

The adjustment of the frequency in response to cheaper car commuting is not uniform among supply regimes, however. The urban and federal operators do increase service frequency, while the suburban and private regimes result in lower f's. The most plausible explanation for this outcome is that the urban and federal governments do consider the welfare loss borne by urban car drivers when congestion intensifies. The suburban and private operators observe a drop in residual rail demand due to the removal of road pricing. For the private monopolist this clearly undermines the profitability of frequency provision. Frequency reduction is more limited in the suburban regimes, as some of the suburban commuters do experience congestion externalities, but overall, lost rail demand and easing of the crowding pressure seems to be a stronger effect compared to the second-best considerations.

In Fig. B.2 we test a wider range of road tax levels, and investigate how the welfare of urban and suburban residents are affected by this parameter. Just like in Table 3 of the main text, the welfare indices are

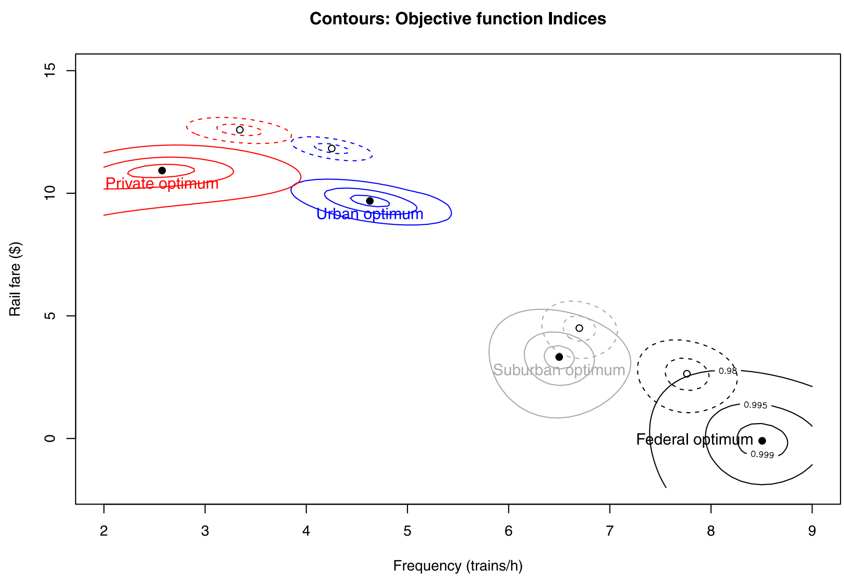

Fig. B.1. Optimal supply levels at $\tau = \$0$ (solid contours), in comparison with the baseline equilibrium of $\tau = \$6$ (dashed contour lines).

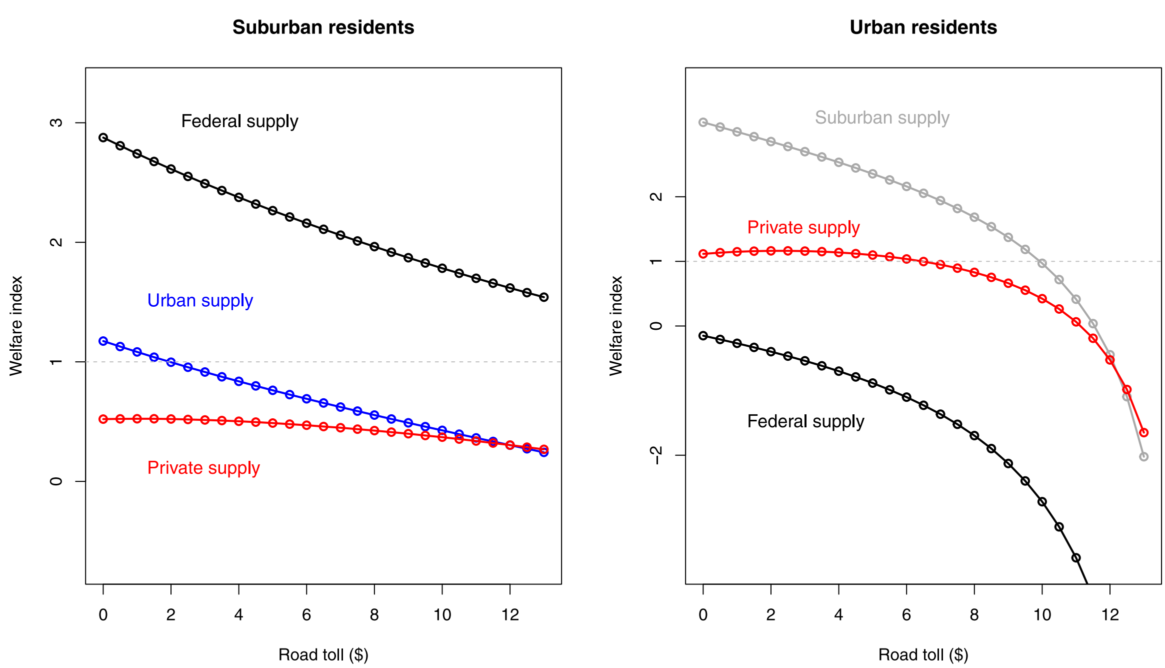

Fig. B.2. Relative welfare outcome of local residents in function of τ, the exogenous road tax level. The horizontal dashed lines represent welfare under local decision making, while the index is zero in the complete absence of the public transport service.

set to 0 in the absence of the public transport service, and 1 corresponds to the optimal supply of each group's own government.[19]

We find the following results. For suburban residents, federal supply remains the most preferred regime in the entire range of tax levels we simulate, and privatisation is always the least preferred option. The attractiveness of the urban government's supply depends, however, on τ. With cheap car commuting, suburban residents would be better off under *both* the federal and the urban regimes. The reason is that excessive congestion puts pressure on the urban government in this case to improve the quality of suburban public transport at its own cost, and suburbanites would benefit from this without contributing to the subsidies.

The right hand panel of Fig. B.2 plots the preferences of urban commuters. From their point of view, federal supply is consistently disadvantageous — even more so when road taxes are high, and therefore suburban public transport does not deliver substantial congestion mitigation. Urban residents prefer private and suburban supply with relatively low tax levels only; the threshold τ value is in the range of \$7 on the 'private' curve, and \$10 on 'suburban' one, under the current set of parameters. With high τ, the residual demand for rail commuting

[19] Fig. B.2 shows that the welfare index decreases with the road tax in both cases. Note, however, that these are *relative* welfare indices, and therefore the *absolute* level of social welfare might show an increasing pattern at certain regions of toll values. With the present parameters, aggregate welfare is maximised at $\tau = \$7$ under federal public transport provision, and \$7.5, \$8 and \$8.5 in the 'suburban', 'urban' and 'private' regimes, respectively.

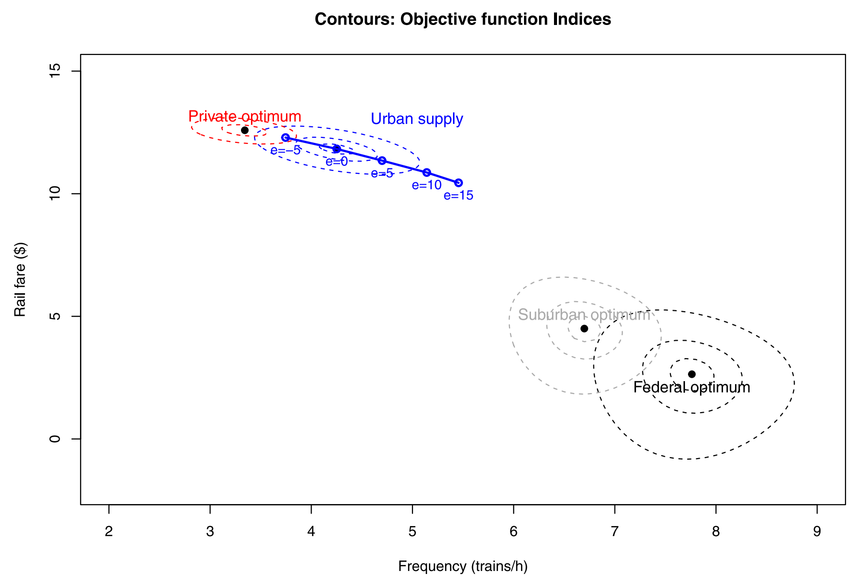

Fig. B.3. Optimal supply by the urban government at various values of the urban road externality. Dashed contour lines represent the baseline equilibrium plotted in Fig. 2.

increases, and thus the urban government becomes capable of using this service for tax exporting purposes under its own control.

Appendix C. Sensitivity test: Urban road externalities

In this exercise we consider that car commuting induces additional external effects in the urban environment, beyond the congestion cost imposed on urban car commuters. Additional urban externalities may include local pollution, accident risks, parking externalities, etc. We test the impact of such effects by adding an additional external cost to the welfare function of urban households. We replace (26) with $W^u(f,p) + Q_c\, e$ in the numerical code, and test the model's sensitivity with respect to the exogenous value of e. Fig. B.3 plots the impact of this parameter on the urban government's optimal fare and frequency offered, and Fig. C.1 shows the welfare effects.

In line with a priori expectations, additional urban road externalities narrow the gap between the optimal supply of decentralised governments. With severe road externalities and limited road taxes, the urban government becomes more interested in inter-regional public transport provision as a second-best measure. However, in the present setup, the value of e must be very high to make rail cheaper and more frequent in the urban regime as compared to the suburban scenario.

Fig. C.1 plots the welfare index of suburban and urban users in the four supply regimes. Once again, this index is 0 in the absence of public transport, and 1 when their own local government provides the service. From the suburban residents' point of view, e has a qualitative impact on how they perceive the urban government's rail supply. When this urban externality is sufficiently high ($e > 9$), suburbanites are better off in the 'urban' regime, as compared to their own government's control. The reason is that in this case the urban government provides a good rail service for suburban commuters to alleviate road traffic, and suburbanites can enjoy benefit spillovers without contributing to subsidies. On the right hand side of Fig. C.1 we see that for urban households, suburban and private supply becomes less attractive as e increases. This is caused by the fact that these types of operators do not take the externality into account in their objectives. However, the attractiveness of the federal regimes improves with parameter e. Severe urban car externalities neutralise one finding in our baseline model, namely that the federal regime is considered a worse outcome for urban residents than having no suburban rail service at all.

Appendix D. Sensitivity test: Alternative service layout

In the core model of this paper, urban residents cannot commute by rail. In many cases suburban public transport service may provide connections within the boundaries of the urban area. The share of urban and suburban residents among users of a suburban public transport service may vary on a wide range, mostly depending on how the administrative borders of the city are drawn and where (and how frequently) stations are located. In this sensitivity test we amend the network layout depicted in Fig. 1 by allowing commuters from Origin #2 (that is, the urban residential area) to commute by rail as well.

In this extension of the numerical model, we assume that urban and suburban residents have identical preferences. To simplify the analysis, let us normalise the travel time on the suburban infrastructure to zero for both modes of transport. These imply that mode split will be identical among the two groups of commuters.[20] Furthermore, we recalibrate road and rail capacities to realise that the congested travel time is twice the uncongested one and rail crowding is 3 passengers per square metre in the initial equilibrium with $x_r^u = x_c^u = x_r^s = x_c^s = 0.25$.

Fig. D.1 plots the welfare indices of suburban and urban residents under the four supply regimes we consider in the paper, varying the share of urban and suburban residents whilst the total population $N = N^u + N^s$ remains constant. The welfare index of suburban (urban) equals unity when the suburban (urban) government is in charge of the service, and drop to zero in absence of public transport.

Fig. D.1 reveals that centralised service provision is strictly preferred by both region when urban residents are also rail commuters, the attractiveness of this regime is even higher for the region in minority. Private supply is the worst option for both groups, independent from the their share in total population. When it comes to decentralised service provision, the distribution of the population has a substantial and mostly trivial impact on the welfare indices. Increasing the share of region A makes service provision by region A more attractive for

[20] With asymmetric preferences and non-identical mode shares, the qualitative results may diverge in various directions. Urban/suburban residents may be in a majority/minority and car/rail users may also be in majority/minority in both groups. A comprehensive analysis of all possible combinations is outside the scope of this paper.

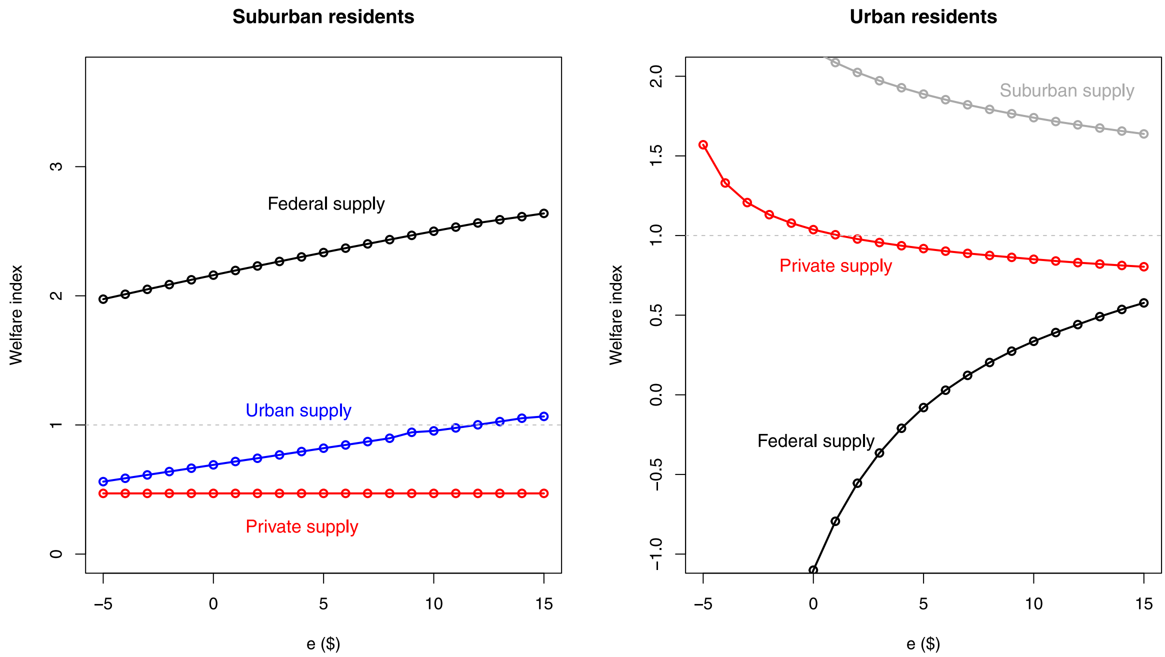

Fig. C.1. Welfare index for local residents in function of e, the urban rail externality.

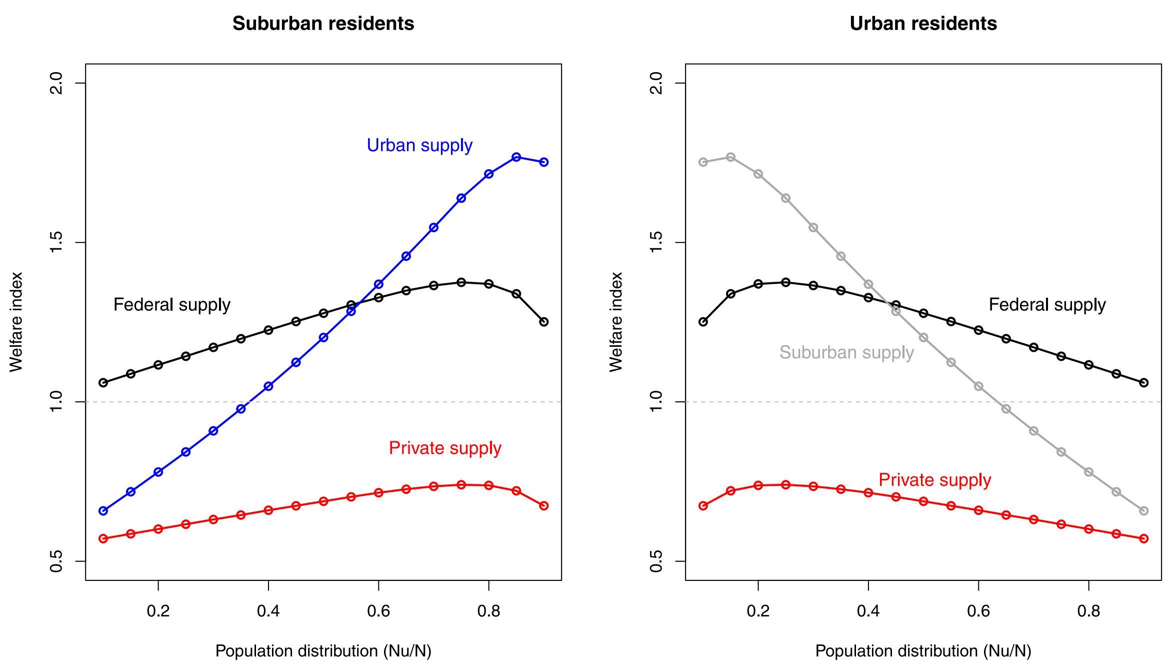

Fig. D.1. Welfare index for local residents in function of N^u/N, i.e., the share of urban residents among all commuters.

the residents of B, in *vice versa*. This implies that, as opposed to our baseline results, the *urban* regime may be preferred by suburbanites as well when N_u/N exceeds a critical threshold. The main reason is that if the service is used by many urbanites, the urban government reduces the fares and increases frequencies, and suburban commuters can exploit this outcome without contributing to subsidies.

The two plots in Fig. D.1 are symmetric due to the assumption that travel times are zero on the suburban infrastructure sections. Therefore, geography has no role in this simplified setting, and the model outcomes are only determined by the distribution of the population. The presence of urban traffic externalities ($e \neq 0$) can be another source of asymmetry in reality. Fig. D.2 repeats the previous simulation with $e = 5$, thus inducing that commuting by car hurts urban residents more.

As we see on the left-hand side, the urban externality has little impact on suburban commuters. The right-hand panel is no longer symmetric, however. Urban inhabitants benefit more from the *federal* regime under which the urban externality is indeed taken into consideration but they can share a part of the subsidy burden with the other region. By contrast, the *suburban* regime is less preferred by urbanites in this case, because the suburban government would neglect the urban road externality in the optimisation of public transport supply.

Future research may consider additional subtleties of the present setting in which urban commuters are also suburban rail users. For example, fares and even service frequencies may be differentiated along the public transport line, thus allowing for price discrimination and targeted tax exporting attempts by decentralised governments.

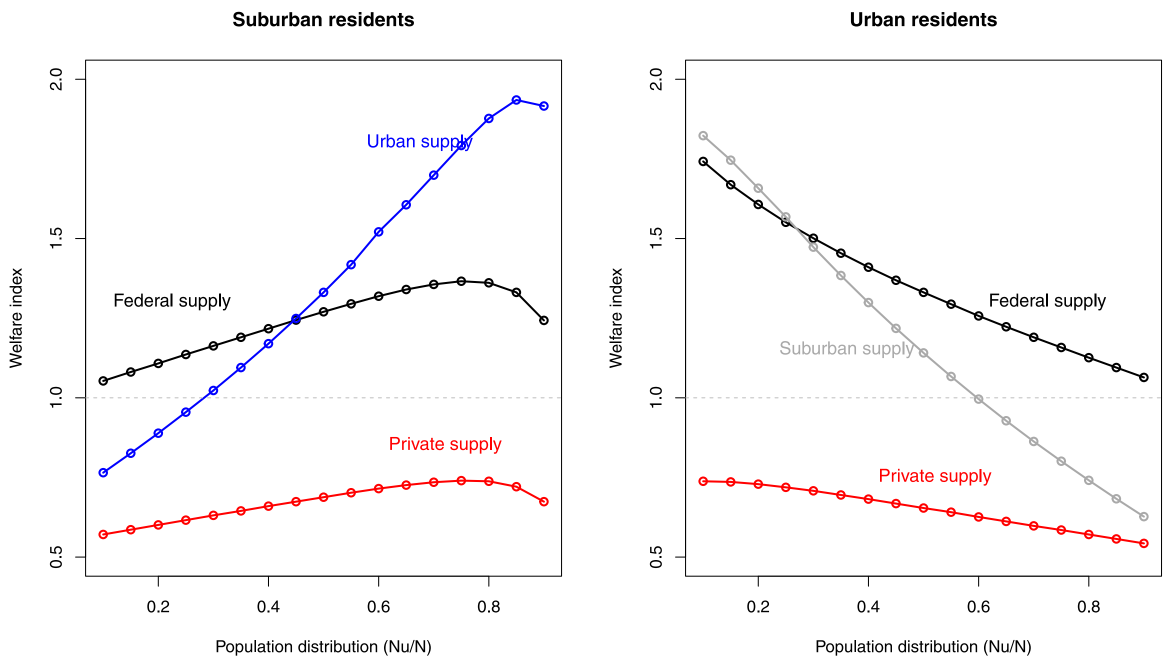

Fig. D.2. Welfare index for local residents in function of N^u/N, with a negative urban externality $e = 5$.

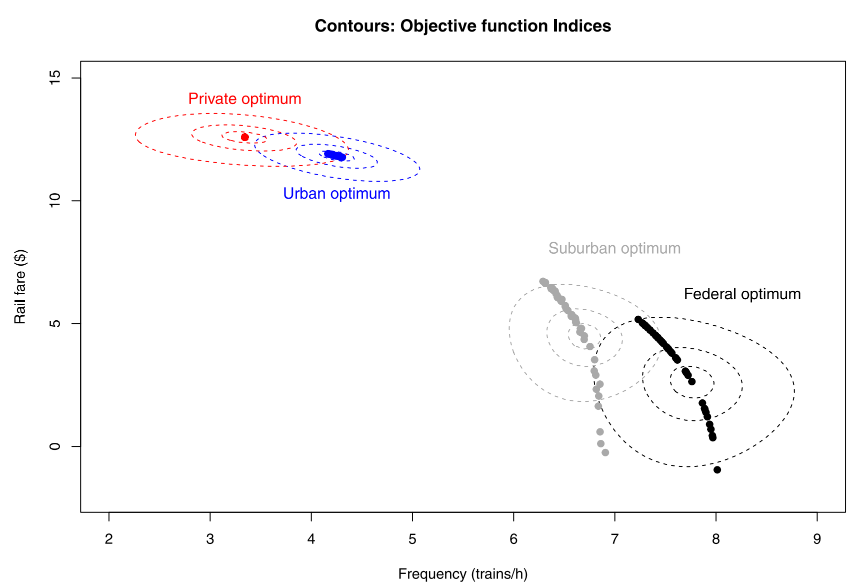

Fig. E.1. Optimal supply with randomly generated parameters of the federal, urban and suburban tax revenue premia.

Appendix E. Sensitivity test: Tax revenue premia

In the baseline version of our simulation model we assume that governments on both levels share the same marginal cost of public funds: $\beta = \beta^u = \beta^s = 1.1$. That is, generating tax revenues to fund the public transport deficit is equally distortionary under all regimes. In practice, the cost of funds for regional governments is likely to strongly depend on the vertical fiscal relations between regional and federal governments, and reliable empirical evidence is not available. In a country with unitary governance, i.e. where the tax levying rights of local governments is effectively controlled by a central government, it would be plausible to assume that local β's are higher than the central one, but there is little empirical evidence supporting this hypothesis.

In the absence of an explicit model describing the fiscal responsibilities and taxing powers of different levels of government and given the lack of more detailed information, we limit the analysis to checking the model's sensitivity with respect to the β parameters. More specifically, without further prior assumptions, we run a randomised experiment in which each government receives a value from a uniform distribution between 1 and 1.2. After each independent draw, we optimise the rail fare and frequency. The results of this experiment are documented in Fig. E.1, where contours of the welfare indices in the baseline equilibrium are marked with dashed lines, and the randomly generated optima are superimposed as a coloured scatter plot.

We observe that the urban optimum shows little sensitivity to the cost of public funds. This is caused by the fact that public transport is close to self-financing in this regime anyway. (Naturally, the private optimum is unaffected by the tax revenue premium.) However, the suburban and federal regimes feature substantial sensitivity along a well-defined path in the fare-frequency plane. In the predefined range

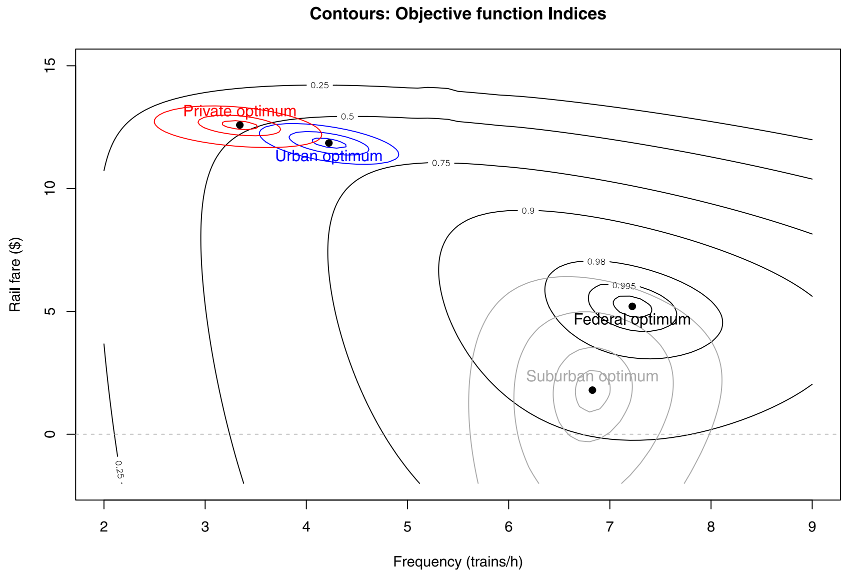

Fig. E.2. Optimal supply with relaxed suburban government budget ($\beta^s = 1.05$) and much higher tax revenue premium on the federal level ($\beta = 1.2$).

Table 5
Policy evaluation with regional welfare indices; $\beta^s = 1.05$, $\beta^u = 1$ and $\beta = 1.2$.

Scenario	Frequency	Fare	Cost recovery	Welfare indices		
	f (trains/h)	p ($)	$Q_r p/z(f)$	Aggregate	Suburban	Urban
Federal	7.2	5.21	0.510	1	1.558	−0.307
Suburban	6.8	1.80	0.186	1.160	1	2.288
Urban	4.2	11.86	0.952	0.637	0.628	1
Private	3.3	12.59	0.966	0.478	0.430	1.048

of the β's, service frequency can never be higher in the suburban scenario than under federal control, but the former may set lower fares, as opposed to what we found in the baseline scenario (see Fig. 2).

To investigate this issue in more depth, Fig. E.2 shows the contour plot of the defined welfare indices for the four governance regimes (equal to one when the own government is in charge) as a function of f and p, by setting $\beta = 1.2$, $\beta^u = 1.1$, $\beta^s = 1.05$. This represent the case when suburban (regional) governments have a greater budgetary flexibility, and therefore the marginal welfare loss of deficit funding is higher on the federal level. We indeed find that suburban governance leads to lower rail fares as compared to the federal regime, but frequencies are still the highest in the latter case. Does this induce a major shift in local governments' ability to pursue a free-rider strategy? To provide an answer to this question, we recalculate the results of Table 3 with the β values above, as shown by Table 5. We find that the ranking of governance regimes from the viewpoint of suburban and urban residents is unaffected by this change in parameters, and the sign of the welfare indices is also robust.

References

Ahn, K., 2009. Road pricing and bus service policies. J. Transp. Econ. Policy 43 (1), 25–53.

Arnott, R., Grieson, R.E., 1981. Optimal fiscal policy for a state or local government. J. Urban Econ. 9 (1), 23–48.

Arnott, R., Yan, A., 2000. The two-mode problem: Second-best pricing and capacity. Rev. Urban Reg. Dev. Stud. 12 (3), 170–199.

Basso, L.J., Navarro, M., Silva, H.E., 2021. Public transport and urban structure. Econ. Transp. 28, 100232.

Basso, L.J., Silva, H.E., 2014. Efficiency and substitutability of transit subsidies and other urban transport policies. Am. Econ. J.: Econ. Policy 6 (4), 1–33.

Besley, T., Coate, S., 2003. Centralized versus decentralized provision of local public goods: a political economy approach. J. Public Econ. 87 (12), 2611–2637.

Boadway, R., Shah, A., 2009. Fiscal Federalism: Principles and Practice of Multiorder Governance. Cambridge University Press.

Borck, R., Wrede, M., 2009. Subsidies for intracity and intercity commuting. J. Urban Econ. 66 (1), 25–32.

Börjesson, M., Fung, C.M., Proost, S., Yan, Z., 2018. Do buses hinder cyclists or is it the other way around? Optimal bus fares, bus stops and cycling tolls. Transp. Res. A 111, 326–346.

Brueckner, J.K., Selod, H., 2006. The political economy of urban transport-system choice. J. Public Econ. 90 (6–7), 983–1005.

Coate, S., 1995. Altruism, the Samaritan's dilemma, and government transfer policy. Am. Econ. Rev. 46–57.

Dahlby, B., 1996. Fiscal externalities and the design of intergovernmental grants. Int. Tax Public Finance 3 (3), 397–412.

De Borger, B., Dunkerley, F., Proost, S., 2007. Strategic investment and pricing decisions in a congested transport corridor. J. Urban Econ. 62 (2), 294–316.

De Borger, B., Proost, S., 2012. Transport policy competition between governments: a selective survey of the literature. Econ. Transp. 1 (1–2), 35–48.

De Borger, B., Proost, S., 2015. The political economy of public transport pricing and supply decisions. Econ. Transp. 4 (1–2), 95–109.

De Borger, B., Proost, S., 2016. The political economy of pricing and capacity decisions for congestible local public goods in a federal state. Int. Tax Public Finance 23 (5), 934–959.

De Borger, B., Proost, S., Van Dender, K., 2005. Congestion and tax competition in a parallel network. Eur. Econ. Rev. 49 (8), 2013–2040.

De Palma, A., Lindsey, R., 2000. Private toll roads: Competition under various ownership regimes. Ann. Reg. Science 34 (1), 13–35.

Friedrichsen, J., König, T., Lausen, T., 2021. Social status concerns and the political economy of publicly provided private goods. Econ. J. 131 (633), 220–246.

Gasparini, L.C., Pinto, S.M., 2006. Equality of opportunity and optimal cash and in-kind policies. J. Public Econ. 90 (1–2), 143–169.

Glaeser, E.L., Ponzetto, G.A., 2018. The political economy of transportation investment. Econ. Transp. 13, 4–26.

Graham, D.J., 2007. Agglomeration, productivity and transport investment. J. Transp. Econ. Policy 41 (3), 317–343.

Graham, D.J., Gibbons, S., 2019. Quantifying wider economic impacts of agglomeration for transport appraisal: existing evidence and future directions. Econ. Transp. 19, 100121.

Hörcher, D., De Borger, B., Seifu, W., Graham, D.J., 2020. Public transport provision under agglomeration economies. Reg. Sci. Urban Econ. 81, 103503.

Hörcher, D., Graham, D.J., 2018. Demand imbalances and multi-period public transport supply. Transp. Res. B 108, 106–126.

Hörcher, D., Tirachini, A., 2021. A review of public transport economics. Econ. Transp. 25, 100196.

Jansson, J.O., 1980. A simple bus line model for optimisation of service frequency and bus size. J. Transp. Econ. Policy 53–80.

Jara-Díaz, S., Gschwender, A., 2003. Towards a general microeconomic model for the operation of public transport. Transp. Rev. 23 (April 2013), 453–469.

Kitchen, H., McMillan, M., Shah, A., 2019. Local Public Finance and Economics. Springer Books.

Knight, B., 2004. Parochial interests and the centralized provision of local public goods: evidence from congressional voting on transportation projects. J. Public Econ. 88 (3–4), 845–866.

Kraus, M., 1991. Discomfort externalities and marginal cost transit fares. J. Urban Econ. 29 (2), 249–259.

Levinson, D.M., 2000. Revenue choice on a serial network. J. Transp. Econ. Policy 69–98.

Li, Z., Hensher, D.A., 2011. Crowding and public transport: A review of willingness to pay evidence and its relevance in project appraisal. Transp. Policy 18 (6), 880–887.

Lockwood, B., 2002. Distributive politics and the costs of centralization. Rev. Econom. Stud. 69 (2), 313–337.

Lörz, O., Willmann, G., 2005. On the endogenous allocation of decision powers in federal structures. J. Urban Econ. 57 (2), 242–257.

Mayeres, I., Proost, S., 1997. Optimal tax and public investment rules for congestion type of externalities. Scand. J. Econ. 99 (2), 261–279.

Melo, P.C., Graham, D.J., Noland, R.B., 2009. A meta-analysis of estimates of urban agglomeration economies. Reg. Sci. Urban Econ. 39 (3), 332–342.

Mohring, H., 1972. Optimization and scale economies in urban bus transportation. Am. Econ. Rev. 62 (4), 591–604.

Mohring, H., 1976. Transportation Economics. Ballinger Publishing Company.

Oates, W.E., 1972. Fiscal Federalism. Edward Elgar Publishing.

Parry, I.W., Bento, A., 2001. Revenue recycling and the welfare effects of road pricing. Scand. J. Econ. 103 (4), 645–671.

Parry, I.W., Small, K.A., 2009. Should urban transit subsidies be reduced? Amer. Econ. Rev. 99 (3), 700–724.

Proost, S., Sen, A., 2006. Urban transport pricing reform with two levels of government: a case study of Brussels. Transp. Policy 13 (2), 127–139.

Small, K.A., Verhoef, E.T., 2007. The Economics of Urban Transportation. Routledge.

Tikoudis, I., Verhoef, E.T., van Ommeren, J.N., 2015. On revenue recycling and the welfare effects of second-best congestion pricing in a monocentric city. J. Urban Econ. 89, 32–47.

Tirachini, A., Hensher, D.A., Rose, J.M., 2013. Crowding in public transport systems: effects on users, operation and implications for the estimation of demand. Transp. Res. A 53, 36–52.

Ubbels, B., Verhoef, E.T., 2008. Governmental competition in road charging and capacity choice. Reg. Sci. Urban Econ. 38 (2), 174–190.

Van Dender, K., 2003. Transport taxes with multiple trip purposes. Scand. J. Econ. 105 (2), 295–310.

Van Dender, K., 2005. Duopoly prices under congested access. J. Reg. Sci. 45 (2), 343–362.

Vandyck, T., Proost, S., 2012. Inefficiencies in regional commuting policy. Pap. Reg. Sci. 91 (3), 659–689.

Wardman, M., Whelan, G., 2011. Twenty years of rail crowding valuation studies: Evidence and lessons from British experience. Transp. Rev. 31, 379–398.

Wilson, J.D., 1999. Theories of tax competition. Natl. Tax J. 269–304.

Xie, F., Levinson, D., 2009. Governance choice on a serial network. Public Choice 141 (1–2), 189–212.