Economics of Transportation 25 (2021) 100196

Contents lists available at ScienceDirect

Economics of Transportation

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

A review of public transport economics

Daniel Hörcher[a,b,*], Alejandro Tirachini[c,d]

[a] Transport Strategy Centre, Department of Civil and Environmental Engineering, Imperial College London, UK
[b] Department of Transport Technology and Economics, Budapest University of Technology and Economics, Hungary
[c] Transport Engineering Division, Civil Engineering Department, Universidad de Chile, Chile
[d] Instituto Sistemas Complejos de Ingeniería, Chile

ARTICLE INFO

Keywords:
Public transport
Public transport demand
Cost functions
Pricing
Capacity provision
Subsidies

ABSTRACT

Public transport provision requires substantial organisational efforts, careful planning, financial contributions from the public, and coordination between millions of passengers and staff members in large systems. Efficient resource allocation is critical in its daily operations. Therefore, public transport has been among the most popular subjects in transport economics since the infancy of this discipline. This paper presents an overview of the literature developed over the past half century, including more than 300 important contributions. With a strong methodological orientation, it collects, classifies, and compares the frequently used analytical modelling techniques, thus providing a cookbook for future research and learning efforts. We discuss key findings on optimal capacity provision, pricing, cost recovery and subsidies, externalities, private operations, public service regulation, and cross-cutting subjects, such as interlinks with urban economics, political economy, and emerging mobility technologies.

1. Introduction

Public transport, defined in this paper as high-capacity vehicle sharing with fixed routes and schedules, is the backbone of urban transport systems in global cities, especially in densely populated metropolitan areas. It is unlikely that mobility will become completely private in the near future, simply because of the inevitable traffic congestion and the difficulties of storing individual vehicles when they are not in use. In other words, even though technological development may transform the appearance of public transport, the fundamental challenge of coordinating between individual travellers who share vehicles of high capacity will remain. The purpose of public transport economics is to make this coordination more efficient, ensuring optimal resource allocation to unlock all societal benefits of mass mobility.

This work reviews more than 300 papers, including the most influential contributions that shaped our understanding of the economics of public transport over recent decades. The earliest studies date back to the 1960s and the 1970s when advanced quantitative methods were not available to calibrate disaggregate supply models, estimate sophisticated demand models, and simulate policy interventions' impact on large urban networks. Did public transport economics significantly change over half a century? Interestingly, the main messages and policy recommendations of the economists in this field are still the same. Scale (density) economies, road pricing, substitution with underpriced car use, socially optimal subsidies, and the peak load problem are still on the research agenda in various forms, just like decades ago. However, the prevalence of popular subjects does not imply that theoretical and empirical results have achieved maximum impact on policymaking. Despite the surrounding consensus among members of the scientific community, the links between scale economies and subsidisation or the limitations of public transport pricing in congestion mitigation are not obvious in the wider transport industry, to mention only two examples. One of the challenges of public transport economics as a sub-discipline will emerge in knowledge dissemination and cross-fertilisation with related disciplines and professions. We believe that a critical overview of past research efforts is crucial in making impactful discoveries in the future.

This paper is not the first review of public transport economics. Many of the pioneering works in the field are reviewed in a book by Nash (1982). Berechman (1993) and Gwilliam (2008) published extensive reviews of the economic and policy issues surrounding public transport, becoming leading sources of information in the context of bus and rail deregulation. The study by Jara-Díaz and Gschwender (2003a,b) is another major contribution in which the authors summarise earlier

---

[*] Corresponding author. Skempton Building, South Kensington Campus, London, SW7 2AZ, UK.
E-mail address: d.horcher@imperial.ac.uk (D. Hörcher).

https://doi.org/10.1016/j.ecotra.2021.100196
Received 18 July 2020; Received in revised form 18 January 2021; Accepted 31 January 2021

developments in the welfare-oriented optimisation of public transport capacity. Mode-specific literature surveys on rail and bus transport were published in the same year by Waters (2007) and Hensher (2007), respectively. Tirachini and Hensher (2012) review the literature of pricing in a multimodal context, where substitution between public transport and underpriced road use is indeed a key aspect. Pricing in public transport was also reviewed by Jara-Díaz and Gschwender (2005) and in a book chapter by Jansson et al. (2015). Finally, there are relevant reviews in closely related disciplines: Desaulniers and Hickman (2007) review optimisation problems in public transport with strong orientation towards operations research, Guihaire and Hao (2008) surveyed papers on network design and scheduling, and Ibarra-Rojas et al. (2015) presented an extensive overview of planning and control problems in bus operations.

The present paper contributes to the literature with a comprehensive review of the microeconomic modelling techniques in the field. In this sense, the paper may serve as a cookbook for future analyses by researchers, students, and professionals. We do not explain the underlying features and mechanisms of each method on a textbook level, but the reader may refer to a large body of literature for such details. This approach also reveals the evolution of methodologies from a historical perspective. The paper primarily covers urban rail and bus travel, but several theoretical insights can be adopted for airborne and waterborne public transport as well. In addition, we present an outlook on emerging modes that share certain features with public transport, including ride-hailing and car-sharing.

The scope of this paper is limited to the welfare economics of optimal policy designs in public transport. Therefore, its orientation is primarily theoretical. The paper reviews relevant empirical findings in the context of model calibration and ex-post policy evaluation (when applicable), but the statistical methodology that such estimates rely on is out of our scope. On the demand side, we discuss ways of representing consumer behaviour in theoretical models and enlist key empirical results suitable for model calibration. We do not cover the literature of public transport user assignment, i.e., models of mode and route choice behaviour in large networks. Similarly, the economic appraisal of long-run investments, such as infrastructure projects, and the cost benefit analysis (CBA) methodology are excluded from the survey. The paper's core subject is the optimisation of supply policies: capacity provision and pricing. We review the evolution of analytical models of optimal frequency, vehicle size, and other supply variables in detail. Pricing and its impact on the degree of self-financing are also investigated. We put public transport supply into a wider context by considering overlaps with the traditional literature of urban economics, industrial organisation, and political economy.

The paper is structured as follows. Section 2 details various potential components of a public transport model, including its demand system, user and operator cost specifications, and how spatial and temporal dynamics are captured. This methodology oriented review is complemented with a classification of the most influential models in the literature, which are presented in the Appendix. Section 3 then turns to the applications of analytical models to various problems of policy optimisation. The majority of the literature considers welfare-oriented supply, perhaps a bit too idealistically. Therefore, Sections 3.1 and 3.7 deal with alternative management objectives and the political economy of public transport, to improve our ability to explain policy decisions in reality. Finally, in Section 4, the review devotes attention to emerging technologies that interact with public transport in its current form and may reshape it in the future. Section 5 presents the study's conclusions.

2. Designing and calibrating public transport models

This section provides an overview and a typology of the most frequently used analytical techniques, highlighting the purpose and basic features of recent methodological contributions. Our discussions are supported by additional references presented in a tabular format in the Appendix. The tables provide a comprehensive overview of the evolution of the literature through the comparison of the methodological toolbox of 38 key contributions in the literature.

- Demand systems are enlisted in Table A.1.
- Table A.2 classifies the papers according to the user cost components discussed in Section 2.3 and the types of temporal and spatial differentiation (Section 2.4).
- Table A.3 details specific technological features of the models and the operator cost functions discussed in Section 2.2.
- Table A.4 presents a range of decision variables in supply optimisation models.

2.1. Demand systems

The fundamental mechanism behind public transport supply decisions is the trade-off between the cost of operations that normally increases with the service provider's output, and mostly travel-time-related costs that users bear in various parts of their journey. User costs normally decrease in the available capacity; for instance, waiting time decreases with service frequency. In the simplest modelling approach, this generic tension can be analysed and optimised by (i) assuming that demand is determined outside the model, (ii) incorporating user cost as part of a social cost function, and (iii) reducing the system optimisation problem into social cost minimisation. In this setting demand enters the model as an exogenous parameter. Social cost can be defined as the sum of operator and user costs, both expressed as a function of the number of users and the capacity variables of interest (e.g. service frequency and vehicle size; see Section 2.2.1). The outcome of such supply optimisation is only applicable in practice if the demand parameter determined outside the model and the optimal capacity derived from the model are in mutual equilibrium. Social cost minimisation leads to the unconstrained (first-best) welfare maximising capacity.[1] An important benefit of the parametric demand approach is that the marginal social cost of a trip, the basis for welfare maximising pricing (see Section 3.3), is simply the derivative of the social cost function with respect to the demand parameter. Thus, in simple settings, this approach enables the derivation of explicit analytical pricing rules for a given level of equilibrium demand, which is often impossible with more complex demand systems.

Replacing parametric demand with a direct or inverse demand function is inevitable when the economic objective behind public transport provision deviates from pure welfare maximisation, to, for example, profit-oriented supply or when a second-best setting is under investigation with pricing or technological constraints. This allows the researcher to quantify the net benefit that consumers attain for service usage and relate it to other elements of the objective function. The sensitivity of demand with respect to the monetary price of travelling determines the supplier's ability to raise revenues by setting fares above the marginal social cost. Demand for public transport can be expressed as a function of generalised travel costs as well, to capture the impact of quality attributes on ridership and consumer surplus. This approach is standard in the general transport economics literature, and widely applied for modelling other (isolated) transport modes (Small and Verhoef, 2007). As a straightforward extension of aggregate models, the demand system can be specified to enable heterogeneity via user-specific parameters in the individual demand function. This creates a suitable framework for modelling price discrimination and non-uniform pricing (see Section 3.3.3).

Mode choice (i.e., substitution between transport modes) is indeed a

---

[1] See Daganzo (2012) for a general discussion on the conditions under which cost minimisation leads to welfare maximising supply, and a public transport specific application in Moccia et al. (2017).

key aspect of many public transport-oriented analyses. In the simplest two-mode setting, public transport and car use can be considered as perfect substitutes. This unrealistic assumption is sometimes made for pedagogical reasons, prescribing that mode split in equilibrium is determined by the equality of generalised user costs in the two modes,[2] but the majority of the literature follows the mechanisms of imperfect substitution via two main paths: (i) Demand and willingness to pay can be derived from a multivariate utility function, or (ii) a discrete choice framework can be established. Both options imply a representative consumer approach in which, at least on the level of predefined groups of travellers,[3] user preferences are homogeneous. Anderson et al. (1992) revealed that the two approaches are actually equivalent under certain conditions.

From a multivariate utility function determined by trip volumes, inverse demand for each mode is derived as the monetary valuation of the marginal trip's incremental utility. The monetary transformation of marginal trip utility is normally performed by adding a numeraire good to the utility function with its price normalised to unity, thus expressing the marginal utility of private income. Alternatively, one may assume a benefit (consumer surplus or *total* willingness to pay function) in monetary terms immediately, in which case the latter transformation can be avoided (see Section 4.5 Small and Verhoef, 2007). If the representative utility function includes interaction terms, for example between the consumption of public transport and car travel, then willingness to pay for one mode will depend on demand for the other mode, thus ensuring imperfect substitution between them. The most usual functional forms for the underlying utility function include the constant elasticity of substitution (CES) and quadratic specifications. The latter is especially convenient for further analytical exercises, as it leads to linear inverse demand functions for each mode (see e.g. Ahn, 2009). Aggregate consumer surplus is expressed in this model as the representative indirect utility multiplied by the number of users. A typical shortcoming arises when the utility functions are quasi-linear, because this assumption eliminates the potentially important income effect when transport expenditure constitutes a substantial share of household income (see Chapter 3 in Jara-Díaz, 2007). Even though this assumption is required to make Marshallian consumer surplus a suitable measure of user benefits, it raises concerns about model adoption in low-income countries.

Besides models of continuous demand variables, discrete choice models are also frequently used in public transport analyses. The majority of this literature follows the tradition of random utility models (McFadden, 1973), with the heterogeneous component of utility assumed to be type-I extreme value distributed; thus, we get logit mode choice probabilities.[4] Both representative utility approaches can be extended to multiple levels of consumer decisions above mode choice, including a distinction between peak and off-peak travel and long-term commitment to car ownership, for example. Such multi-level models are evaluated recursively: Utility associated with alternatives on higher levels are assessed based on the expected surplus of choice situations on lower levels. Small and Rosen (1981) derive that expected utility in the choice situation can be transformed into the traditional monetary measure of consumer surplus by normalising it with respect to the marginal utility of income. With a logit specification, expected utility boils down to the frequently used *logsum* formula. This convenient property is exploited in a numerical appraisal of competing multimodal urban transport policies by Basso and Silva (2014), Tirachini et al. (2014b), and Hörcher and Graham (2020b), among the most recent contributions. The practical downside of using discrete choice demand systems, especially in their simplest multinomial logit form, is their inflexibility during calibration; it is difficult to replicate any combination of own and cross demand elasticities drawn from empirical exercises.

As one moves from relatively simple, aggregate representations of space towards real networks, additional discrete travel decisions have to be considered on the demand side, including route choice. Given the simultaneous dependency between demand and user costs on network segments, reaching an equilibrium requires public transport assignment.[5] Network modelling implies that the demand system has to be disaggregated to the level of a representative user for each spatially differentiated origin-destination market, or at least to arrival and alighting rates at stops (Toledo et al., 2010). With advances in computational power, further disaggregation is made possible. Agent-based demand systems handle a population of synthetic travellers individually. This way heterogeneous user characteristics and preferences can be modelled very precisely. Dedicated software widely used in the academic community for network-level public transport modelling include MATSim (Horni et al., 2016), MILATRAS (Wahba and Shalaby, 2005), and BusMezzo (Cats, 2013). In multi-agent demand systems, aggregate behaviour is recovered from the simulation of individual decisions during travelling. Links exist with activity based models that include decisions before and after individual trips, thus reproducing entire daily trip chains (Bekhor et al., 2011). This creates ground for demand prediction at a very high resolution at the expense of increased efforts in data collection, parameter calibration, and the derivation of system equilibria. Without sufficient empirical evidence in the calibration process, disaggregate models may do more harm than good. However, the increasing availability of high-resolution demand and flow data due to the massification of low-cost Information and Communication Technologies (mobility apps, traffic counts) eases the process of calibrating agent-based models in large areas.

Large-scale agent-based models are rarely used in traditional economic analyses due to the lack of transparency in the relationship between aggregate variables and because of the difficulties of deriving general results from a model calibrated for a specific city or a geographical area. However, certain elements of agent-based modelling have the potential to be adopted in public transport economics in a more simplified network configuration due to the inherent benefits of this approach in reproducing demand heterogeneity. For example, MATSim has been used to optimise bus headway and fare (Kaddoura et al., 2014, 2015).

The calibration of a demand model requires data collection from the specific geographical area of interest for direct parameter estimation, or the researcher may rely on measurements of demand sensitivities published in the literature. An easily applicable measure of demand sensitivity is its elasticity with respect to key travel attributes such as fare level, service quality, journey time components, income and car ownership, and price of competing modes (Oum et al., 1992). Hundreds of elasticity estimates are available in individual studies, review articles and meta analyses, including more recent contributions by Paulley et al. (2006), Wardman (2012), and Wardman (2014). As an rule of thumb and international average, Paulley et al. (2006) propose that the price elasticity of bus demand is −0.4 in the short run (1–2 years), −0.56 in

---

[2] The original Downs–Thomson paradox is one of the typical examples of such multimodal setups governed by the equality of equilibrium user costs (Mogridge, 1997; Basso and Jara-Díaz, 2012; Zhang et al., 2014). Note that this approach is equivalent to Wardrop's principles, a concept widely used for modelling route choice in a road network where perfect substitution is indeed much more plausible than in a two-mode problem.

[3] The number of sub-groups of representative users could be increased substantially with the advent of high speed computing. In the extreme case, each household of a geographic area can be modelled as an individual decision-maker, which leads us to the emerging literature of agent-based models of public transport supply.

[4] Exceptions include the linearisation of the logit function (Kocur and Hendrickson, 1982) and a uniformly distributed idiosyncratic taste parameter in Basso et al. (2011a).

[5] Assignment is out of the core scope of this paper; the interested reader is referred to a substantial body of literature reviewed by Liu et al. (2010) and Gentile et al. (2016a).

the medium run (5–7 years), and −1.0 in the long run (12–15 years). Urban rail price elasticities are −0.3 in the short run and −0.6 in the long run. A meta-analysis of Holmgren (2007) finds that for U.S. cities, short-run demand elasticities with respect to the level of service, income, price of petrol and car ownership are 1.05, −0.62, 0.4 and −1.48, respectively. The empirical literature provides evidence of user preferences for different modes within public transport, and transfers between them in large networks (Hensher and Golob, 2008; Varela et al., 2018; Garcia-Martinez et al., 2018).

Note, however, that demand elasticities are very context specific. Some regularities have been identified as part of the reviews above. For example, elasticities with respect to price are higher in rural areas than in metropolitan regions; peak demand is less sensitive to tariffs than off-peak demand; the demand elasticity might increase with income, demand for leisure trips is more elastic than work or school related trips, and larger fare deviations in the empirical setup normally lead to greater elasticities. This implies that a careful selection of baseline elasticities is essential for successful model calibration. Authors usually deal with less reliable elasticity parameters by performing sensitivity analysis to assess the degree of robustness of models to changes in such parameters. It has to be emphasised that elasticities are not unique parameters of the demand model; in most cases, they vary along the demand curve considered. Thus, careful calibration might not end with the reproduction of a baseline equilibrium, but the researcher has to be confident that the demand model remains realistic even if larger deviations in supply are considered in the analysis. Unfortunately, the empirical results in the literature are also point estimates around the observed equilibria, which makes it difficult to validate numerical models along a wider range of demand levels. In this respect, a comparative evaluation of the demand systems reviewed earlier in the present section is an outstanding task on the research agenda.

2.2. *Transport operations*[6]

Economic models of public transport require an adequate representation of the underlying technological process. The capacity made available for passengers, which is an *intermediate output* of the transport operator, is the relevant outcome of the technological process of public transport service provision. This phase of service provision can be modelled with traditional microeconomic tools: Technology determines the production function of input factors, and under standard conditions the mix of inputs is optimised for a given level of intermediate output (measured, e.g., in vehicle kilometres) such that the cost of production remains minimal. Consequently, capacity as an intermediate output will then become an important determinant of operator costs, on the one hand, and the quality of service as perceived by the user, on the other hand. Capacity imposes an upper bound on the quantity of the *final output*, the number of passengers transported. In Section 2.2.1 we first review the most frequently used dimensions of capacity in public transport, and then Section 2.2.2 describes the operator cost functions considered in the empirical and theoretical literature. Technology may affect user costs through several ways, for instance through boarding and alighting times, and the impact of information provision on waiting time valuation. Thus, the present discussion has a direct link to Section 2.3, where user costs are determined by the available capacity, among other system characteristics.

2.2.1. *Public transport capacity*

The engineering interpretation of public transport capacity is the maximum number of passengers that can be transported along a route, given the supplier's intermediate outputs, such as service frequency and vehicle size. In this interpretation, capacity is usually measured as the hourly flow of passengers. One may also distinguish the capacity of *vehicles* from line capacity. The latter comes as the product of hourly frequency and the capacity of vehicles. Service frequency is constrained by a number of technological and design variables. In the case of buses running on segregated bus lanes, bus stops generally have lower capacity than signalised intersections. Therefore the number of buses circulating is constrained by the capacity of the bus stops, which must have sufficient space for buses to queue (Fernández and Planzer, 2002). The throughput of bus stops is determined by the demand level and by several engineering decisions such as (i) the number of berths and the possibility of overtaking at bus stops, (ii) the bus length, (iii) the number and width of bus doors, (iv) the passenger boarding policy (if boarding is allowed only at one door or at all doors), (v) the fare collection technology and (vi) the number of passengers boarding and alighting (Gibson et al., 1989; Tirachini, 2014). On the other hand, in mixed operations where cars interact with buses, a large car flow may congest signalised intersections or make the access of buses to bus stops difficult; therefore, cars may indeed heavily restrict bus flow levels and capacity. In the case of rail systems, maximum train flow is constrained by the minimum safety headway enabled by the signalling system.

The maximum number of passengers per vehicle is affected not only by engineering variables, such as the number of seats and the area provided for standing (if allowed), but also by social aspects such as the level of occupancy that is accepted within vehicles. While no more than 3 or 4 passengers per square metre are acceptable in some countries, 6, 8 or 10 passengers per square metre are allowed in other countries, particularly in busy metro lines (Basu and Hunt, 2012; Tirachini et al., 2013), generating extremely uncomfortable travel conditions.

The transport economics literature uses the term *capacity* in a broader context; it may cover a range of variables that capture the technological characteristics of the public transport service. Beyond frequency and vehicle size already mentioned in the engineering context, this may also include the number of seats inside the vehicle, the number and size of doors, the number of stops, and the route length. One of the main goals of public transport economics is to develop supply rules to optimise the capacity variables, pursuing a predefined objective. Capacity variables will have important roles in economic models even if demand remains below the physical capacity.

In microeconomic models of public transport, researchers often assume that capacity variables are *responsive* to marginal changes in other model variables, such as the level of demand. In practice, the assumption of responsiveness means that bus or train operators are able to readjust frequency and vehicle size after marginal changes in demand. Is this assumption realistic under any circumstances? It is only realistic (i) in the planning phase of new services, (ii) when a large fleet of vehicles is available for the operator so that vehicles can be quickly reassigned between routes, or (iii) if the operator has access to a (secondary) market for public transport vehicles where capacity can be purchased or sold relatively quickly. Moreover, the operator might be unable to react to any increase in demand in the short term if the peak-hour frequency of services is already at the maximum possible, having to resort to long-term solutions (e.g., creation of new lines and infrastructure investments). Bus and rail based services might differ in terms of capacity responsiveness. As the fixed infrastructure cost of bus service expansion is lower, and the vehicles themselves are cheaper, bus operators can usually react more quickly to demand shifts.

The responsiveness of capacity has a significant impact on the cost structure of public transport use. If, on the level of the intermediate output, there is an underlying rule that determines the optimal level of capacity in function of demand, then the incremental user will trigger a marginal capacity adjustment and, consequently, a deviation in operational and user costs as well. The incremental trip's operating cost is lower if capacity is fixed. Some of the user externalities are also linked to responsive capacity. Most importantly, density economies *in user costs*

---

[6] Table A.3 in the Appendix classifies major contributions in the literature according to the technological details they include.

due to the frequency dependency of waiting time disappear if frequency can no longer be increased on the margin (see Section 3.2), leading to higher fares and lower subsidies in optimum (see Sections 3.3 and 3.4 for more details of the Mohring effect). In summary, microeconomic models of optimal capacity outputs are suitable to uncover general relationships between variables of interest, such as demand, service frequency, and the social cost of travelling; they do not necessarily represent the operational regime of perfect capacity adjustment.

Capacity imposes an upper bound on ridership. By normalising passenger flows by service frequency we see that vehicle capacity must be relevant in this respect. The literature is divided in how vehicle capacity and occupancy (vehicle load) are modelled. Some papers include an explicit restriction on the maximum number of passengers on board, for example, Jansson (1980), Oldfield and Bly (1988), and Basso et al. (2011a).[7] A common feature of this approach is that, in the absence of crowding costs, the capacity constraint is always binding under cost minimisation, because having spare capacity involves extra operator costs without providing benefits to users. Thus, the optimal vehicle size is directly derived as the ratio of the passenger load at the most demanded section of a route and optimal frequency. Other authors simply assume that the capacity constraint is never binding (this is often the case in dynamic models; de Palma and Lindsey, 2001) or that the operator maintains a predefined occupancy rate under varying demand conditions (e.g., De Borger et al., 1996). The explicit capacity constraint can be replaced with crowding dependent user cost specifications. In this case, it is the inconvenience of crowding that keeps demand below the physical limits of vehicle capacity (e.g. Jara-Díaz and Gschwender, 2003a,b; Pels and Verhoef, 2007). This approach can be extended with the user cost of failed or denied boarding and queueing before boarding. Denied boarding as an equilibrium outcome requires a dynamic model framework that Kraus and Yoshida (2002) proposed, among others.

2.2.2. Operator cost functions

It is the researcher's natural desire that the representation of transport operations in a supply-side model should be based on sound empirical evidence. The estimation of cost functions has a long history in transport economics, given that cost estimation is a highly relevant input for policy evaluation. Some policy questions can be mentioned here. First, microeconomic theory suggests that a public monopoly has to be subsidised under optimal pricing and scale economies, that is, when the marginal social cost of production is below its average cost. Second, in a competitive environment, scale economies can lead to natural monopolies in the absence of regulation, so cost functions provide key industry structure indices in dialogues surrounding deregulation. Third, the accuracy of payment rules from authorities to public transport operators depends on having reliable and transparent cost functions.

However, defining the cost function and the appropriate measures of scale economies is not a straightforward task in case of a transport operator. Comprehensive reviews of this subject are presented in Jara Díaz (1982), Oum and Waters (1996), Braeutigam (1999), Jara-Díaz (2007), Pels and Rietveld (2007), and Basso et al. (2011b). Selecting a measure of output is a key decision point in cost function estimation. Purely technical efficiency-oriented analyses are normally based on the intermediate output (e.g., the vehicle, car, or seat kilometres produced). In fact, public transport provision contracts are usually based on delivering a predefined transport capacity set in this way. Some authors argue that the final output of transport, that is, the actual passenger flows or passenger-kilometres, reflects the ultimate economic motive behind service provision (De Borger et al., 2002) and is the appropriate measure of the effectiveness of the transport provider's service offerings (Small and Verhoef, 2007).

Another focus point where the literature is yet to achieve consensus is how scale economies[8] in transport firms are quantified. It is commonly agreed that the spatial dimension of production is a distinctive feature of this industry. A significant part of the literature follows Caves et al. (1984) who define *returns to density* as the impact of total capacity on average cost keeping network size constant, and propose that *returns to scale* is the impact of proportional changes in traffic density and network size (Oum and Waters, 1996; Pels and Rietveld, 2007). Another stream in the literature, hallmarked by Jara-Díaz and Cortés (1996), Jara-Díaz and Basso (2003), and Basso and Jara-Díaz (2006), proposes a refinement of this approach based on multiproduct firm theory. These studies argue that returns to density resembles more what is normally considered returns to scale for multiproduct firms; that is, the impact of equi-proportionate changes in OD-level flows keeping the number of OD pairs served constant, and the calculation of this metric in real applications should consider the local degree of homogeneity with respect to disaggregated flows (Jara-Díaz and Cortés, 1996). Also, the authors propose the use of economies of spatial scope instead of economies with respect to network size because the economies of spatial scope consider each origin-destination (OD) pair in a network as a separate product of the firms.[9] At the intermediate level, they point out that beside the choice of network *size*, measured by the number of points served, line structure is another important decision in the production process. Cost functions with constant line structure should be distinguished from those with variable network layout. In summary, it is apparent from the discussion above that crucial policy questions related to subsidisation and deregulation can no longer be answered with the estimation of a unique parameter of returns to scale, due to the spatial dimension of transport activities and cost interdependencies in networks.

In terms of frequently used cost function specifications, Pels and Rietveld (2007) split the literature into two groups: basic functional forms, such as the Cobb-Douglas, Leontief, and CES specifications, and more flexible alternatives, including the translog, generalised Leontief, quadratic, and generalised McFadden specifications. The translog cost function is indeed the most widely used specification in the literature. However, the econometric identification methodology has advanced since the translog specification became popular. Most studies apply simple least-squares estimation techniques, which require that all explanatory variables, including output and input prices, are exogenous. This assumption is indeed questionable (Basso et al., 2011b), for example, due to the monopsony power of large transport operators on the labour market. Another potential source of endogeneity is that productivity/inefficiency might be an unobserved factor of the firm's production function, leading to omitted variable bias in the estimation (see Sections 2.2 and 2.3 of Anupriya et al., 2020a, for a deeper elaboration). Savage (1997) addressed endogeneity with an instrumental variables approach, but the choice of instruments is a highly context-dependent challenge that cannot be generalised for any applications. In a more recent study, Anupriya et al. (2020a) applied dynamic panel generalised method of moments estimation to control for confounding from both observed and unobserved covariates; they found that returns to density and network size are underestimated with traditional econometric tools in a case study of the global metro industry.

In urban bus and rail transport, the presence of increasing returns to density (at a constant network size) is a consensual empirical finding (Pels and Rietveld, 2007; Basso et al., 2011b), which makes the threat of natural monopolies and the need for public subsidies under optimal

---

[7] See other papers indicated in the "vehicle capacity constraint" column of Table A.3.

[8] Microeconomic theory defines returns to scale as a property of production functions, while scale economies characterise cost functions. Under standard assumptions widely adopted in the transport economics literature, these properties are interchangeable, but not necessarily under any assumptions.

[9] Pels and Rietveld (2007) and Batarce (2016) acknowledged the theoretical merits of the multiproduct approach but pointed out that its empirical application is very demanding due to the need for disaggregate output data.

pricing undeniable. The picture is more diverse when it comes to returns to network size or spatial scope: The literature can be split into estimates of neutral and positive returns to spatial coverage. In general, rail networks are more prone to the emergence of natural monopolies due to large fixed costs, especially when the network size is limited and, therefore, complex interchange stations are not needed (see Anupriya et al., 2020a, for empirical evidence from the metro industry).

Cost functions create associations between some aggregate measures of output and the cost of operations. The theoretical foundation behind most of the empirical work is the theory of cost-minimising firm behaviour, that is, the combination of inputs to produce a given level of output, subject to input prices. Surprisingly, little is utilised from what we know about the specific supply-side considerations of public transport operators. In Section 3, we review how key design variables, such as service frequency and vehicle size, are set in light of a management objective, given the demand characteristics the firm is facing. We will see that a given amount of aggregate output, measured in vehicle or passenger kilometres, for example, can be produced with a wide range of frequency and vehicle size combinations, leading to a wide range of operator (and user) costs. Such considerations are absent from the current literature of cost function estimation. This implies that calibrating operator cost functions for supply optimisation is not a straightforward task either as directly applicable empirical estimates are not available from the literature.

The literature relies on several types of operator cost specifications in supply optimisation problems.[10] Many studies assume that ridership has a direct impact on operator costs. This makes the derivation of marginal operator costs easy. However, it is more realistic to assume that demand has an indirect impact on operator costs through the capacity policy applied. That is, the direct sources of such costs are the number and size of vehicles. The former can be represented with either the frequency of services or the fleet size, which are equivalent if the cycle time is exogenous. The choice of vehicle size is indeed an important decision where the technological efficiency of larger vehicles can be exploited and transformed into density economies. Frequency and vehicle size are the most frequently considered determinants of operator costs. The multiproduct nature of public transport is sometimes recognised in supply-side models through temporal and spatial cost interactions, which are discussed in more detail in Section 2.4.

2.3. *User cost functions*[11]

Public transport does not differ from other transport modes in the sense that certain costs of displacement are borne by the users themselves. In elementary transport economics, it is the cost of time loss that adds to the monetary price of travelling. In public transport, the actual time lost when travelling inside the vehicle is just a fraction of the total door-to-door travel time. This section reviews the most important elements of supply models from a user cost perspective, including the state-of-the-art empirical estimates.

2.3.1. *Trip stages and user cost components*

The journey experience in public transport can be split into a number of distinct stages. Each phase has its own duration, and the opportunity cost of foregone time may also differ in each stage. The literature distinguishes the following journey time components:

- Access time: time required to access the first stop or station where the vehicle boarding takes place, from the trip origin.
- Waiting time: time spent waiting at a bus stop or station platform before boarding a public transport vehicle.
- In-vehicle time: time spent aboard a vehicle while travelling towards a destination.
- Transfers: when using more than one public transport vehicle to perform a trip, a transfer is required, which apart from involving walking and waiting time, is a source of inconvenience by itself due to the disruption of the trip as perceived by the user.
- Egress time: time from the alighting of a vehicle at the final stop or station to the actual destination of the trip.

Note that multiple types of user costs may belong to each journey segment above, especially to the in-vehicle experience, including extra costs due to travel discomfort, unexpected delays and travel time variability. The standard assumption in the literature is that user costs in individual trip stages are independent from each, and these costs can be summed up to get the *generalised* cost of the trip. This generalised cost is expressed in either temporal or monetary dimension. In the former case, *travel time multipliers* are defined for each trip stage, in order to transform the magnitude of inconvenience into the time loss that causes the same disutility for the consumer. Alternatively, user costs can be expressed in monetary units as well, by applying separate *values of time* for trip segments. The value of time is the amount of money that a passenger is willing to pay to reduce his/her travel time by one unit. In the majority of analyses, these values of time are exogenous parameters estimated in dedicated empirical exercises. As opposed to the theory of the opportunity cost of pure time loss (see, e.g., Becker, 1965; DeSerpa, 1971; Jara-Díaz, 2007), user costs associated with walking, waiting or travelling in crowding do not have widely accepted micro-foundations.

2.3.2. *Access and egress time*

Krygsman et al. (2004) described access and egress time as “the weakest links in a public transport chain that determine the availability and convenience of public transport”. They estimate using a comprehensive travel-activity diary where access and egress times constitute 20%–50% of the total trip time for most multimodal trips in the Netherlands. Brons et al. (2009) claimed that improving and expanding access services to railway stations can substitute for improving and expanding the rail service itself, and the former is often more cost-effective than the latter. Krygsman et al. (2004) added that “accurate estimates of access and egress times are often in short supply”, despite their huge impact on the attractiveness of public transport.

Public transport can be accessed by multiple modes of transport. It is a legitimate question how we should distinguish public transport trips with high access cost from multimodal trip chains in which public transport is one element of the chain. The standard assumption in the literature is that bus stops or rail stations are accessed on foot. In this case, the key input variables of access cost calculation are (i) the walking distance, and (ii) the unit cost of walk time. The walking distance is jointly determined by the spatial distribution of trip origins and destinations and the location of access points for public transport. The latter plays an important role in the optimisation of stop spacing and route density (see Section 3.2.4). Item (ii), or the factors of the value of walking time, is considered in a simplistic way in the literature. The inconvenience of walking might be affected by a number of external factors, such as the weather, as well as the quality of the built environment. In addition, the perceived security is an important factor that is far too often neglected in relation to public transport use in all stages of a trip, and the empirical results indicate clear gender differences in this respect (Delbosc and Currie, 2012; Börjesson, 2012; McCarthy et al., 2016; Ait Bihi Ouali et al., 2020).

Nevertheless, authors of supply studies normally just pick one value of walk time estimate from the empirical literature (see e.g., Wardman, 2004; Paulley et al., 2006) to calibrate the cost of system access. One potential reason behind this simplistic approach is the belief that transport operators have limited control over the walking part of the

---

[10] Table A.3 reviews the typical operator cost specifications among key publications.

[11] Table A.2 in the Appendix classifies major contributions in the literature according to their user cost components.

journey experience. Indeed, rail systems with spaced stations are also commonly accessed by motorised modes. In a wider policy context, a more sophisticated modelling of the access experience may shed light on an important source of seemingly exogenous shocks in demand (e.g., higher access costs in bad weather) and validate the claim of Brons et al. (2009) when they stated that cost-effective access cost reduction has a great potential in making public transport more attractive.

2.3.3. *Waiting time and scheduling delay costs*

The chief driver of passenger waiting time is the headway between vehicles, that is, the time interval between the arrival of two consecutive vehicles at any given stop or station. Mean headway and its variability are crucial to determine the amount of time passengers have to wait. The literature defines service frequency as the number of vehicles providing commercial service in any given point of a route, per unit of time. The most common unit measure for service frequency is the hourly number of departures. By construction, headway is the inverse of service frequency, that is, the duration between the arrival of two consecutive services.

The relationship between frequency and the actual waiting time depends on more complex behavioural assumptions. When service frequency is relatively high, the modeller may assume that passengers have no chance to adjust their activities to scheduled public transport departures, so they arrive at the stations randomly, at a constant rate. In this case, in theory, it is not important to define and communicate an exact timetable of vehicle movements, because users' scheduling behaviour remains random anyway. If the headway between vehicles is fixed and constant, the luckiest user's arrival coincides with the vehicle's departure, and therefore she does not encounter delay due to waiting. By contrast, the longest waiting time under regular conditions is the headway itself. Thus, the expected wait time is assumed to be half of the headway in this case. With random arrivals, waiting is inconvenient for three main reasons:

(i) The passenger loses time, which may be spent on leisure or other productive activities.
(ii) Depending on the quality of the stop or station infrastructure, waiting might be inconvenient due to the physical effort of standing, weather conditions, and station crowding, among others.
(iii) The presence of waiting time brings inevitable uncertainty regarding the trip arrival time, which is costly for users who might have a preferred arrival time.[12]

The primary behavioural impact of relatively long headways is that the expected waiting time increases. At some point, consulting a timetable and adjusting activities prior to the trip to a planned departure may become more attractive for passengers than enduring long waiting times at stations. In this case, there is a waiting time cost outside stations because departures are not at a user desired time. Without an explicit preferred departure time pattern (see Footnote 12), the expected user cost of activity rescheduling can also be modelled as a penalty that increases with the headway between vehicles (Tirachini et al., 2010b). As *passive* waiting enables the passenger to undertake productive activities or leisure, its opportunity cost is expected to be lower than the value of *active* waiting time loss experienced at stops or stations. In other words, parts of items (i) and (ii) in our list above are eliminated if passengers arrive to the stop or station non-randomly.

It is indeed an important research objective to determine the critical headway where passenger behaviour transitions between random arrivals and timetable-based trip planning, in order to understand the wider relationship between the value of headway and service frequency. The critical headway is normally estimated to be between 5 and 12 min (Fan and Machemehl, 2009; Ingvardson et al., 2018; Berggren et al., 2019). The actual value depends on several context dependent factors. For example, the quality of station infrastructure may shift the critical headway upwards due to convenient waiting conditions. On the other hand, with the advent of online vehicle location information systems and real time travel time prediction, access to reliable information becomes cheaper, shifting the critical headway downwards (De Borger and Fosgerau, 2012). In this case, a reduction on the value of waiting time savings is also achievable due to a better use of waiting time and a reduction of anxiety caused by more reliable timetables. We observe a tendency that more recent estimates of the threshold headway are generally lower than in earlier studies. Through a semiparametric regression framework Singh et al. (2020) found that the representative passenger of the London Underground shift to non-random arrivals above less 3 min, depending on the time of day. Nevertheless, as Ingvardson et al. (2018) and Singh et al. (2020) documented, substantial heterogeneity exists among passengers in terms of their trip planning practices.

So far, we considered a direct relationship between headways and the *planned* service frequencies. In fact, headways are always subject to certain variability, which is especially prominent along bus routes that share the right-of-way with other vehicles, through the well-known phenomenon of bus bunching. For illustration, let us assume that we have a service with an average frequency of 10 vehicles per hour; if headways are regular, the average waiting time is 3 min. However, if buses arrive in groups of two vehicles every 12 min, the average waiting time doubles to 6 min. Therefore, for the same level of aggregate frequency, headway variability has doubled the average waiting time. In reality, a number of additional factors contribute to bunching; these factors include congestion, demand shocks, and differences in boarding or alighting times. Osuna and Newell (1972) and Newell (1982) estimated average waiting time models as linear functions of the variance of headways.[13]

Finally, in high-demand periods, waiting time increases due to denied boardings, i.e. when some passengers are unable to board a vehicle due to overcrowding. Failed boarding lengthens the waiting time by an additional headway. In this case, the expected bus or train waiting time might be formulated as a function not only of frequency but also of the occupancy factor of vehicles, defined as the ratio between the number of passengers inside a vehicle and its total capacity (e.g., Oldfield and Bly, 1988; Cepeda et al., 2006).

2.3.4. *In-vehicle time and crowding*

In supply-side models, the duration of the in-vehicle travel time is a critical determinant of the optimal policy outcomes. In-vehicle travel time might be uniform or heterogeneous among passengers, depending on the spatial layout and temporal dynamics of the model (see Section 2.4 for the related discussion). Travel times may be endogenous functions of the supplier's decision variables, for example due to congestion effects or the duration of boarding and alighting. Thus, in-vehicle travel time functions will be key ingredients of the determination of optimal capacity and pricing in Sections 3.2 and 3.3.

---

[12] Fosgerau (2009) and Fosgerau and Engelson (2011) derive the marginal social cost of headway adjustment and the user cost of the resulting travel time variance, based on standard assumptions on passengers' scheduling preferences. Schedule delay is defined as the deviation between the passenger's preferred departure or arrival time and the realised trip schedule. The main consequence of the random waiting time due to the headway between services is that the passenger cannot avoid schedule delay with certainty. The schedule delay cost further increases if headways themselves are uncertain.

[13] The bunching effect can be reduced by slowing down certain vehicles at the expense of the in-vehicle travel time loss of passengers travelling on these vehicles. Optimal control algorithms for dispatching are outside the scope of this review (see e.g., Muñoz et al., 2013; Berrebi et al., 2015; Gkiotsalitis and Van Berkum, 2020).

The value of in-vehicle time savings is affected by a range of nuisance factors beyond the pure opportunity cost of lost travel time (Haywood et al., 2017). Technological features of the vehicle, such as its seating configuration, are natural determinants of the trip quality and thus the value of time (Wardman and Murphy, 2015). Crowding refers to the state of high vehicle occupancy in which users experience disutility due to the lack of personal space and limitations in moving around inside the vehicle. The discomfort caused by physical proximity between passengers on board had gained particular attention in the literature due to the wide range of impacts that the crowding externality has on both users and operators (see in-depth reviews by Tirachini et al., 2013; Hörcher, 2018).

Crowding cost estimation is a rapidly growing branch of the demand modelling literature. Wardman and Whelan (2011) and Li and Hensher (2011) reviewed the empirical evidence gathered until the early 2010s. Since then, new contributions follow the consensual approach in which the cost of crowding is identified as a multiplier of the value of in-vehicle travel time, as opposed to earlier experiments with additively separable in-vehicle time and crowding cost elements in consumer utility. Researchers tend to agree also in the way how vehicle occupancy rates are measured. Dropping the previous practice of relating demand to the vehicle's seating capacity, recent studies measure the density of passengers on a unit of in-vehicle area, to make their estimates independent of the seat configuration. Along these lines, the literature has expanded with new estimates based on revealed preference data (Kroes et al., 2014; Tirachini et al., 2016; Hörcher et al., 2017), and stated preference studies revealing heterogeneity in crowding valuations (Tirachini et al., 2017; Yap et al., 2020) and testing non-parametric heterogeneity distributions (Bansal et al., 2019). A parallel branch of the empirical literature identifies crowding as an important determinant of passenger satisfaction, especially for peak-hour bus services (Allen et al., 2019; Börjesson and Rubensson, 2019).

2.3.5. Transfers

Transfers are essential to exploit density economies on high capacity arteries while maintaining connectivity in a large network. Transfers are certainly costly for passengers, partly due to the loss of travel time, but also because disruptions in the travel process is perceived as an inconvenience in itself. Nevertheless, transfer costs are not frequently considered in supply models of public transport primarily because the spatial scope of such analyses is often restricted to representative origin-destination pairs or a single line.[14] For the purposes of demand modelling and especially for transit assignment, the literature delivers a series of empirical estimates of the value of transfers. These estimates are often split into three components: the valuations of (i) walk time during transfers, (ii) additional wait time, and (iii) the pure inconvenience of disrupting the continuity of a trip, often called as the *transfer penalty* (see, e.g., Garcia-Martinez et al., 2018).

Currie (2005) reviewed the empirical literature and finds that the transfer penalty depends heavily on the public transport modes involved. The in-vehicle travel time equivalent of the penalty remains under 15 min in studies of urban rail transfers, but it reaches the region of 30–50 min when it comes to bus transfers. He argues that the major force behind this result is the quality of interchange infrastructure. Raveau et al. (2014) identified additional factors that influence the transfer penalty in the metro networks of London and Santiago. They show, controlling for the time loss of transferring, that the vertical distance between platforms, the direction of movements and the availability of escalators or elevators are significant determinants of the transfer penalty. This hints that a fixed penalty is not necessarily a good representation of the transfer experience; a more detailed set of walk and wait time multipliers may capture more of what is often believed to be a fixed user cost (Iseki and Taylor, 2009).

How much of this empirical evidence is utilised in supply models? Given that the literature is gradually moving towards more disaggregate representations of public transport networks (see Section 3.2.4), it is likely that the importance of transfer valuations will also increase in the future.[15] Fielbaum et al. (2018) reacted to the large variability reported in the literature by commenting that "further research on this is badly required". We agree with this claim because transfer quality may affect several economic network characteristics as well. For example, it is likely that in dense urban rail networks, transfer costs increase with the complexity of station infrastructure, and this may have a detrimental effect on returns to network size.

2.3.6. Modal specificities

Even though most of the key messages of public transport economics are not mode-specific, certain properties of the models do have to reflect the transport technology considered. One can distinguish the bus and rail-oriented groups of analyses in the literature. Bus-oriented models pay more attention to various disturbances caused by interactions with traffic and the complexity of boarding and alighting. Thus, cycle time in most bus models is endogenously determined. Naturally, buses may contribute to (and are usually affected by) road congestion. The net contribution of bus usage to congestion also depends on the degree of substitution between private car and bus demand. The delay cost of traffic interactions opens up the room for the establishment of dedicated bus lanes; several papers, including Basso et al. (2011a); Basso and Silva (2014) and Börjesson et al. (2017), consider this as an explicit decision variable in supply optimisation.

Other important features of bus transport relate to the design of stops. The technological processes of entering and leaving the bus stop and allowing passengers to board and alight are additional sources of delays. First of all, the capacity of bus stop itself (in terms of the number of vehicles they can serve simultaneously) may limit the degree of density economies in service frequency. Second, the duration of bus dwelling depends on the number of passengers who wish to board and alight the vehicle in a particular stop. The relatively low investment cost of bus stops (except for BRT stations) implies that route density and stop spacing are more often considered as demand responsive variables in bus models. Both of these variables affect the access cost of public transport, that is, the time required to walk to the closest stop, and the literature shows that similar returns to scale exist in this user cost as what drives the Mohring effect in case of waiting time (see Section 3.2.4).

Boarding and alighting technology is among the main differences between bus and rail operations: The number and size of doors normally increases with the length (size) of trains. Thus, assuming that train length is responsive, the duration of boarding and alighting may remain independent of the representative demand level that rail services face (Harris, 2006; Harris and Anderson, 2007). From a modelling point of view this implies that the cycle time is also independent of demand conditions, and therefore boarding and alighting passengers do not impose a travel time externality on those already on board. Although under very crowded rail operations it might happen that travel time increases with demand, due to crowding-induced delays at stations (Tirachini et al., 2013).

An obvious feature of densely used urban rail systems is the need for large stations, often under ground. This adds a considerable amount of walking time to the full price of travelling. That is, the share of waiting and in-vehicle travelling may be lower in case of rail modes, which implies that some of the supply-side features of public transport derived from the cost of waiting (e.g., the Mohring effect) may be weaker than what we find for buses. On the other hand, movements inside large

---

[14] See Section 2.4 and Table A.2 for our summary.

[15] In an early contribution, de Jonge and Teunter (2013) showed that an appropriate treatment of the walking phase of transfers leads to substantial improvements in bus timetable planning.

stations, especially during peak periods, may generate frictions between simultaneous users and produce externalities in economic terms. A typical example is queueing at fare gates when entering or exiting stations, which is a typical dynamic congestion problem where deadweight loss is generated without appropriate incentives.[16]

2.3.7. User equilibrium

Public transport models follow the mainstream transport economics literature when the behaviour of users is described (Small and Verhoef, 2007). User equilibrium is reached when no traveller is able to attain further personal benefits by adjusting her trip volume (on a continuous scale) or switching to other travel alternatives (in case of discrete user decisions, e.g., route or departure time choice). This requires that the indifferent marginal user's benefit from travelling equals to its personal *generalised price*. Depending on the demand system considered, the marginal user benefit can be (i) expressed as an explicit inverse demand function or (ii) derived by the differentiation of a utility function with respect to individual travel demand. The generalised price includes all user cost components and monetary payments associated with the marginal trip, all transformed and aggregated in a monetary dimension. As in public transport models the number of user cost components can be numerous, and, as discussed above, capacity on the supply side might also be responsive, finding a system equilibrium numerically is often a more challenging task compared to traditional road traffic models. Thus, analytical solutions to the user equilibrium problem with elastic demand are rare in the literature, but state-of-the-art computing techniques in transit assignment enable numerical solutions even for very large networks (Gentile et al., 2016b).

2.4. Spatial and temporal dynamics in networks[17]

The spatial and temporal configuration of public transport provision determines many cost and demand interdependencies between markets served by the multiproduct supplier. Networks are represented in various ways in the literature. The simplest but still widely used modelling approach is to neglect such interdependencies and reduce the spatial and temporal pattern into a representative OD pair. This enables the fundamental mechanisms behind supply optimisation to be analysed in a transparent way, in many cases analytically. In an aggregate model all demand and supply side variables have to be considered as an *average* of their network-wide distribution. The cost of transparency is indeed the averaging bias. The geographical and temporal distribution of the production process (again, we refer back to multiproduct firm theory; Jara-Díaz and Cortés, 1996; Basso and Jara-Díaz, 2006) and heterogeneity in demand conditions between spatio-temporally differentiated markets (Hörcher and Graham, 2018, 2020a) may cause significant distortions in model outputs for a given level of *average* network condition.

An intermediate solution between aggregate and fully disaggregate modelling is the application of correction factors that capture the impact of demand and cost interdependencies in the real network. Bly and Oldfield (1986) provided an early example. In their aggregate model the average waiting time cost is expressed in function of the average occupancy rate. However, they consider that due to spatial and temporal demand imbalances, the capacity constraint will first become binding in the most popular markets, increasing the probability of failed boarding and excessive waiting. Thus, the average waiting time depends on the *distribution* of occupancy rates as well. Oldfield and Bly (1988) discussed several potential functional forms for the waiting cost function, all featuring an adjustment factor imposed on the *average* occupancy rate to take the peakiness of demand into account. Unfortunately, these factors are always empirically calibrated, without relating them to a more general measure of the degree of demand imbalances. More recently, Moccia et al. (2017) adopted a similar approach to express crowding costs in function of the average occupancy rate.

Arguably, adding more time periods or network segments to a supply model allows for a more transparent representation of unbalanced demand as well as cost interdependencies. The two most popular approaches are (i) differentiating peak and off-peak regimes in the model and (ii) defining a network with more than one line segment to represent spatial heterogeneity; the simplest back-haul setting has two such segments on a bidirectional link.[18]

The two-period and bidirectional setups can be extended into a complete line (e.g., Rietveld and van Woudenberg, 2007; Tirachini et al., 2014b), or a more general multiperiod framework with spatial and temporal demand fluctuations and fixed capacity (Pels and Verhoef, 2007; Hörcher and Graham, 2018). For specific types of analyses, the core direction of travel along a corridor can be defined *a priori*, thus creating a link to the frequently used monocentric city model of urban economics (e.g., Kraus, 1991; Basso and Silva, 2019). Considering that both dimensions of the urban space come into picture when walking to and from public transport stops or stations, the spatial density of lines and stations can also be modelled endogenously. The majority of such studies assume that demand is generated uniformly across a predefined space, and the role of public transport is to connect this demand to one point in space, i.e. an urban activity centre or larger transport hub (Kocur and Hendrickson, 1982; Small, 2004). When the choice between network configurations is also part of the design problem, and transfer costs are included among user costs, then urban space is represented by a set of nodes with the possibility of linking them multiple ways, depending on network structure (Gschwender et al., 2016). Fielbaum et al. (2016) proposed that typical urban forms (monocentric, polycentric and dispersed cities) can be represented with no more than three variables of a parametric city model. Finally, the highest granularity is achieved with comprehensive models of the urban road and rail network (Toledo et al., 2010), often combined with an agent-based demand system. These grant reliability for practical applications but prevent access to transparent analytical insights.

3. Optimal supply policies

The purpose of the demand and supply side models reviewed in the previous section is in most cases to analyse optimal public transport provision, i.e. to optimise pricing and capacity variables with respect to a predefined economic objective. This section covers this subject of major policy importance via the description of potential policy objectives in Section 3.1, the derivation of welfare maximising capacity (Section 3.2) and pricing (Section 3.3) rules, and their implications on public subsidies (Section 3.4). Section 3.5 extends the scope of our analysis to supply regulation in liberalised markets, while Sections 3.6 and 3.7 review the role of this mode in the spatial economics and political economy literatures.

3.1. Policy objectives

The first step of supply optimisation is indeed to define an objective function that the planner intends to maximise or minimise. The vast majority of the studies we review considers either social welfare or profit maximisation, or the second-best problem of welfare maximisation subject to a budget restriction. These choices follow the traditions of welfare economics, and the notion of social surplus and profits are well

---

[16] This property can be observed in the summary of Table A.3, where we see that rail models, as indicated in Table A.1, normally do not feature endogenous cycle time.

[17] Table A.2 in the Appendix classifies major contributions in the literature according to the type of temporal and spatial differentiation of service provision they analyse.

[18] Table A.2 documents several examples from the literature.

understood in the narrow field of transport economics. However, one might have the impression that these objectives are restrictive when it comes to industrial practice. The policy impact of research in transport economics is often hindered by the mere fact that the concept of social welfare is not well understood or accepted in the policy arena. Therefore it is a natural desire from the viewpoint of the academic community to showcase the benefits of welfare-oriented planning with simpler performance metrics, or, alternatively, to come up with other objectives that are easier to interpret for the wider public but resemble the outcomes of welfare maximisation. We return to such alternative objectives after reviewing the more traditional and widely applied approaches of welfare and profit maximisation.

The concept of social surplus, defined as the sum of costs and benefits perceived by users, suppliers and third parties in society, is the driving principle behind the literature of optimal public transport supply. Social welfare is the sum of net user and supplier benefits, extended by further welfare effects on other members of society, when applicable. The way how consumer benefits are modelled is strongly dependent of the underlying demand specification. We saw that in the simplest case of parametric demand, the welfare function boils down to the minimisation of various social cost components (e.g.,the sum of user and operator costs). Glaister (1974) introduced the framework of *generalised consumer surplus* in which *willingness to pay* covers user expenditures in multiple dimensions, including time and inconvenience. If demand is derived from an underlying utility function in a representative consumer framework, then aggregate consumer surplus is simply the product of indirect utility in user equilibrium and the number of representative users. Under the logit specification, consumer surplus is expressed as the expected utility in the choice situation, measured by the *logsum* formula (Small and Rosen, 1981), and normalised by the marginal utility of money. Thus, random utility theory can be mobilised to assess the economic efficiency of supply-side policies as well.

As soon as an appropriate demand system is established, profit maximisation is straightforward from a modelling point of view: revenues can be captured as the product of demand and the prevailing fare, and the sum of revenues and negative operator costs yield the net financial result of service provision. Although pure profit maximisation does provide some analytical insights, it rarely appears as the sole objective of supply optimisation in public transport. In this industry, monopolies are normally publicly owned or regulated, for example, if private competition is not feasible due to the threat of emerging natural monopolies. Financially constrained welfare maximisation turns out to be more relevant for policy analysis. The literature shows that depending on the tightness of the budget constraint, the resulting supply behaviour may vary between unconstrained welfare maximisation and profit maximisation. Specific second-best policy outcomes under a financial constraint are discussed in Section 3.3.2.

Interestingly, alternative objectives beside social welfare and profit functions were more frequently discussed in the late seventies and early eighteens than today. Nash (1978, 1982) investigates the maximisation of passenger miles and vehicle miles subject to a budget constraint.[19] Demand and vehicle mile maximisation can be achieved with a profit-oriented objective as well, if the monopolist receives a subsidy after each unit of output. He shows that this subsidy equals the inverse of the Lagrange multiplier of the budget constraint in the Lagrangian function of passenger miles or vehicle miles maximisation. In a numerical exercise, Nash (1978, 1982) found that ridership maximisation subject to a zero profit constraint is just slightly less efficient than welfare maximisation under the same constraint. By contrast, vehicle miles maximising supply might cause even greater harm in economic terms than the monopolistic exploitation with profit maximisation.[20]

Glaister and Collings (1978) investigated ridership maximisation in combination with various weighting schemes. Predefined weights can be assigned to individuals or groups of users in the objective function. They conclude that with an appropriately designed system of weights, passenger miles maximisation can indeed converge to welfare maximising supply. Although *weighted passenger miles* may sound easier to interpret for a non-economist audience, Glaister and Collings (1978) warned that finding the optimal weighting system requires as much information on demand as computing the classical social welfare function. Based on a separate analysis, Bös (1978) added that the maximisation of passenger miles does not lead to unambiguous redistribution benefits either. Frankena (1983) showed that one cannot tell whether ridership maximisation leads to fare and frequency levels above or below the second-best welfare optima, without knowing the actual demand and cost functions. Moreover, he concludes that the impact of subsidies to a firm which maximises ridership is also ambiguous.

From the above-cited pioneers of public transport economics, we learn that no easily interpretable objective function can replace welfare maximisation, which may explain why the literature had subsequently evolved along the lines of classical welfare economics. Nonetheless, this orientation is not consensual in the public transport *industry*. Nash (1978) noted more than four decades ago that the objective of Pareto-type social welfare maximisation was “too complicated to command widespread understanding and support” among industry stakeholders, and this observation may hold today equally well. Therefore, investigating supply under alternative objective functions, in various policy scenarios, can help the researcher understand the reasons behind sub-optimal decisions and assess their economic consequences.

3.2. *Optimal capacity provision*[21]

The public transport economics literature has gone a long way providing insights into optimal public transport service and capacity levels, particularly through the development of microeconomic models that abstract to some extent from real-life complexities, to elicit general insights by finding both analytically and numerically the optimal value of decision variables such as service frequency, the size of vehicles, station or stop density and network structure.

Despite the technological complexities of advanced capacity models, the underlying economic rationale behind optimal supply has remained persistent over recent decades. The general rule states that capacity should be increased as long as its marginal benefit is greater than its marginal cost, no matter which mode and what capacity variables we consider. Passengers normally realise benefits from higher capacity. Depending on the exact formulation of the optimisation problem’s objective, marginal benefits may include elements of consumer surplus as well as reductions in user costs. On the other hand, the total operator cost naturally increases with the service capacity provided. Thus, the tension between benefits and costs leads to the optimisation problem at the heart of microeconomic models of capacity. Developments in the literature are linked in most cases to the introduction of new capacity

---

[19] Indeed, both are realistic alternatives in the political environment in which public transport services are provided. Vehicle miles maximisation can be communicated as an attempts to make as much capacity available as possible within the financial restriction, while demand maximisation seeks to make the service attractive for as many consumers (voters) as possible.

[20] Nash (1978, 1982) considered another management objective as well: revenue maximisation subject to a zero-profit constraint. In this case, fares are very high, but the revenues are reinvested into excessive vehicle miles provision, so that profits in the end fall to zero. Even though common-sense intuition may suggest that this objective is less harmful than the monopolist’s, Nash finds that under his set of parameters profit maximisation is still more desirable. From a policy point of view, this implies that the absence of profits does not guarantee economic efficiency, if fares are set too high.

[21] Table A.4 in the Appendix classifies major contributions in the literature according to the decision variables they consider in supply optimisation.

variables, improvements in the demand system, or restrictions in the optimisation problem in the form of financial or technological constraints.

3.2.1. *Frequency*

The cradle of this body of literature was established by Mohring (1972), who attempted to find the optimal value of the service frequency for a single bus route. Mohring's model is built on the trade-off between users' waiting time cost and the operator cost of service frequency. The sum of these two cost components is minimised when the marginal waiting time benefit of frequency adjustment equals its marginal operator cost. Under three assumptions, this optimality condition leads to a very simple frequency rule.

(i) The crucial assumption is that total waiting time cost must be linear in both the number of users (Q) and the headway between services (f^{-1}), such that it takes the form $Q{\cdot}kf^{-1}$, where k is a constant that includes the value of waiting time. This condition is satisfied with random passenger arrivals and regular headways; see Section 2.3.3.
(ii) The operator's cost is linear in frequency. This implies neutral scale economies in the production of this intermediate output.
(iii) The third assumption is that frequency has no further impact on any parts of the service provider's objective function.

Under assumptions (i) and (ii), the marginal waiting time cost reduction is $-Q{\cdot}kf^{-2}$, and the marginal operator cost is just a constant that we denote here with ρ. After setting the sum of these two marginal cost components equal to zero to achieve the optimum, Mohring derives after rearrangement that the optimal frequency must be proportional to the square root of demand: $f=\sqrt{Q{\cdot}k/\rho}$. This rule is often mentioned in the literature as "Mohring's square root principle".

The concave dependency between demand and the optimal frequency has a number of highly relevant economic consequences. Most importantly, it guarantees increasing returns to the scale of the *final* output (ridership), even though in assumption (ii) the model assumes neutral scale economies in the *intermediate* output. This mechanism is often called as the *Mohring effect*. As demand grows, both the average waiting time and average operator cost decrease, and therefore the marginal social cost of public transport trips remains below the average cost. This is a major argument supporting fares below the average operator cost, and thus loss generating operations supported by public subsidies. We discuss more details of the model's pricing and cost recovery consequences in Sections 3.3.1 and 3.4.

This basic optimisation approach has been subsequently enriched to include a series of other effects that frequency may have on the technological process or the user's experience. For example, Jansson (1980) considered that if cycle time depends on demand per vehicle through boarding and alighting times, then frequency determines the duration of in-vehicle time as well. Kraus (1991); Jara-Díaz and Gschwender (2003a,b), and Tirachini et al. (2014b) included the discomfort of passenger crowding that increases the value of in-vehicle time savings. Jansson (1980) and Jara-Díaz et al. (2017) considered that frequency can be differentiated between peak and off-peak operations but the fixed cost linked to fleet size is unaffected by off-peak demand conditions. Tirachini and Hensher (2011) added bus congestion to the model, in the form of queueing delays at bus stops. Leiva et al. (2010) introduced spatial differentiation in frequency through fleet assignment strategies such that express services can skip certain stops. Delle Site and Filippi (1998) and Cortés et al. (2011) analysed the optimal frequency with short-turning and deadheading. Finally, service frequency remains an important part of capacity optimisation in combination with vehicle size, network structure and stop spacing that are treated in the coming sections. In all models in which waiting time appears in the objective function and assumption (i) above is maintained, the first-order condition of optimal frequency can normally be rearranged to resemble Mohring's original square-root formula.

3.2.2. *Beyond frequency: Optimal scheduling*

Beside the frequency of a public transport service, the timing of each departure may also matter for the consumer. This leads us to the timetabling problem. Newell (1971) investigated waiting time minimising dispatching for a smooth distribution of exogenous passenger arrival rates over time. It is more realistic to consider, however, that passengers might have preferred departure and/or arrival times and perceive schedule delay costs associated with early and late travel. Scheduling preferences and the timetabling problem resemble the classical models of location decisions in spatial economics, and, more generally, optimal supply with product differentiation.[22] If preferred travel times are distributed over a closed time interval (e.g., one day), without substitution between days, then timetabling is equivalent to Hotelling's line model. By contrast, Salop's circle model is applied when preferences vary around the clock, the transport provider sets the time of hourly departures, and travellers might reschedule outside a one-hour interval as well.

Alfa and Chen (1995) studied the properties of user equilibrium in the line model, considering bus capacity constraints and failed boarding as well. Assuming homogeneous distribution of desired arrival times, de Palma and Lindsey (2001) showed (with no capacity constraint) that the optimal timetable leads to regular headways in both the line and circle models. They derive additional analytical results with heterogeneous schedule delay costs. In the Hotelling model, departures are more widely spaced in this case, meaning that the first bus departs earlier and the last bus departs later. This is a form of product differentiation aimed at accommodating passengers with strong aversion to schedule delay; de Palma and Lindsey (2001) urged that neglecting heterogeneity leads to a sub-optimal timetable and higher than optimal average user cost. Later on, de Palma et al. (2015) introduced crowding costs and differentiated pricing into the problem. They find that the optimal timetable is unaffected by crowding if scheduling preferences are homogeneous. However, with a peak in the preferred arrival time pattern, the optimal train departures are more narrowly spread around the peak, and differentiated pricing can further improve efficiency by diverting demand away from the most crowded service(s). In Section 3.3, we return to the dynamic trip scheduling framework with a review of pricing studies in which timetables are no longer endogenous.

3.2.3. *Vehicle size*

Vehicle size optimisation, intrinsically linked to frequency optimisation, becomes an important subject of public transport economics in the 1980's. Several studies contributed to the "minibus debate" initiated by Walters (1982), who claimed that density economies in Mohring (1972) and Nash (1978) mostly disappear when one allows for endogenous bus size. Walters found that high-frequency minibus services could outperform the ones operated with standard 55-seat urban buses, and the gap between marginal and average social costs vanishes in the former case, together with the justification for subsidies. In his response, Mohring (1983) confirmed that minibus services might operate under milder density economies. He pointed out, however, that Walters' results are mainly driven by the assumption that minibus drivers' wage rates can be equivalent to taxi drivers', which is just one third of average bus driver salary. Mohring (1983) found that even if this difference in resource costs was achievable in a competitive minibus market, the

---

[22] This section covers *temporal* product differentiation in public transport. Variety can also be introduced in other attributes of service provision: Gronau (2000) addressed the vehicle size problem in the framework of product differentiation. He proposes that if a market is sufficiently thick and passengers' value of time is dispersed, then multiple parallel public transport services with unequal vehicle size could improve efficiency.

minimum average social cost is just marginally different from standard bus operations. Gwilliam et al. (1985b) expressed stronger criticism, stating that keeping the average waiting time with standard buses independent of demand is “a simple error in Walters’ model”. They show, after replacing this assumption with the waiting time set equal to half the headway, that the optimal bus size increases with demand (as opposed to what Walters derives). Gwilliam et al. (1985b) found that numerical optimisation of the correct model with the parameters that Walters (1982) used yield “something like a standard U.S. single decker”. Debates about the optimal bus size became highly policy relevant in Britain, as support for bus deregulation, reviewed in Section 3.5 of this paper, was centred around the expectation that private operators would run smaller buses more frequently (Glaister, 1986).

In the simplest static models reviewed so far, conclusions can be drawn about the optimal vehicle size by imposing a capacity constraint on the model, that is, assuming that the ratio of ridership and the optimal frequency is constant.[23] Jansson (1980) was the first author to consider separate peak and off-peak operations, where the vehicle capacity constraint is expected to remain binding only in the peak. He shows analytically that frequency and waiting time becomes less important, relative to bus size, as fewer passengers are carried outside the critical sections and time periods of service provision. In other words, an uneven demand pattern results in larger optimal bus size. Oldfield and Bly (1988) returned to the single-period static model of a representative OD-pair. On the other hand, they make advances in the way how the average waiting time is modelled, allowing for the possibility of failed boarding when buses are too small and demand fluctuates. They investigate three waiting time assumptions:

(i) Constant average load factor, which creates a deterministic relationship between demand, the frequency, and vehicle size.
(ii) A load factor dependent *multiplier* is imposed on the waiting time function so that wait time depends on both service frequency and the occupancy rate, because the expected cost of failed boarding increases with vehicle occupancy.
(iii) A load-factor-dependent cost is *added* to the standard headway dependent waiting time cost.

Oldfield and Bly (1988) were also innovative in the sense that they calibrated their model with elastic demand. They found after calibration that assumption (i) leads to the highest optimal vehicle size, while assumption (ii) generates the lowest. Exogenous subsidies have a positive impact on vehicle size in all three cases.

The literature reviewed so far assumes that user costs are independent of the occupancy rate, and so operators have no interest in increasing vehicle size, unless the overall capacity constraint becomes binding. In an empirical analysis of train lengths in the Netherlands, Rietveld et al. (2002) found that occupancy rates are systematically below 100%, and suggest that “future theoretical work in the field of choice of vehicle size and frequency should pay explicit attention to various reasons” behind it.[24] Jara-Díaz and Gschwender (2003a,b) were the first to include the dependency of the value of in-vehicle travel time on the occupancy rate in an extension of Jansson (1980). The timing of this development in the theoretical literature coincides with the appearance of growing empirical evidence on the disutility of crowding, often measured as a multiplier on the value of time (see Section 2.3.4). Jara-Díaz and Gschwender (2003a,b) showed in a numerical exercise that this extension has a considerable positive impact on the optimal frequency as well, compared to the original models of Mohring and Jansson. Pels and Verhoef (2007) derived first-order conditions for the optimal vehicle size in a more general demand system with imperfect modal substitution multiple origin-destination pairs along a line.[25]

Basso et al. (2011a) introduced a feature that may be specific to bus services: They considered that the manoeuvrability of buses depends on their size, and define the time in motion in function of vehicle size. Still in the context of buses, Tirachini et al. (2014b) assumed that vehicle size as a decision variable can take four discrete values only, in correspondence with what is commercially available: These are the mini, standard, rigid long and articulated configurations. Jara-Díaz et al. (2020) proposed that two fleets of vehicles of different size should be used to tackle demand fluctuations between peak and off-peak periods.

Tirachini et al. (2014b) introduced the number of seats as an additional decision variable, thus making the actual capacity of vehicles more flexible than what the four discrete bus types would imply. They find in their numerical exercise calibrated with data from Sydney that maximising seat supply is optimal choice when crowding emerges. Hörcher et al. (2018b) derived an analytical rule for optimal seat provision in a general framework, which suggests that lower than maximal seat supply may be optimal under different parameters. In a rail context, de Palma et al. (2015) keep vehicle size constant, but allow seat provision to vary. They do so in a dynamic peak-period setup in which crowding discomfort on the busiest trains can be regulated by reduced seat provision. Indeed, reallocating space between seated and standing users may improve passenger capacity on the busiest trains. Later on, de Palma et al. (2017) performed the reverse analysis with fixed seat layout and endogenous train length, testing various pricing regimes in the dynamic peak setup. They reveal that in this specific setup, capacity investments and efficient pricing are not substitutes for congestion relief, that is, capacity expansion delivers higher benefits when other supply variables are set to social optimum.

3.2.4. Route and network structure

Stop spacing, line spacing and network design determine the spatial properties of public transport provision. Early contributions including Mohring (1972) have already documented the key trade-offs involved in decisions about the spatial density of services. Supply density has a negative impact on the user cost of accessing the system (normally by walk) and a positive one on in-vehicle travel times and various operator costs.

In general, access time could be decomposed into access to the route, that is roughly perpendicular to the route that the user wants to take, plus access along the route to the stop. The former depends on the distance between routes (route density) whereas the latter is given by the distance between stops or stations (stop density). Kocur and Hendrickson (1982) and Chang and Schonfeld (1991) modelled this behaviour in the simplest way by assuming that demand is uniformly distributed over space, there is a set of parallel routes uniformly separated from each other, and all are equally attractive to the users. With this approach, the uniform distance between routes is a decision variable. The maximum distance that a user will walk is half of the route spacing, and the minimum distance is zero (the origin of the trip is along one of the routes). The average walking distance to the route is then one quarter of the line spacing. To estimate average walking time along the route, an analogous procedure is performed. The average walking

---

[23] See a series of papers applying an explicit vehicle capacity constraint in Table A.3.

[24] This statement was not entirely correct, as the discomfort of standing (Kraus, 1991; Hörcher et al., 2018b) as well as the probability of failed boarding (Oldfield and Bly, 1988) could already explain why train capacities were not fully utilised.

[25] Pels and Verhoef (2007) showed that welfare and profit oriented operators share the same vehicle size rule, in line with a more conventional result from studies on private roads (Small and Verhoef, 2007, see Section 6.1.1). They explain that the monopolists’ incentive to supply socially optimal vehicle size is that any reduction in user costs through lower crowding turns into higher revenues for the operator. However, as the monopolist charges higher fares, the equilibrium vehicle size is unlikely to remain identical to what the welfare oriented operator sets.

distance with uniform demand distribution and stop density turns out to be one quarter of the stop spacing. Tirachini et al. (2010a) applied this framework to a radial public transport network in which passengers access radial arteries through tangential streets. Other authors assume a limited set of possible locations for bus stops or train stations determined by road intersections and other characteristics of the built environment (Furth and Rahbee, 2000; Chien and Qin, 2004). Geographic Information System (GIS) tools are suitable to identify more precise walking distances to stops and can be embedded in discrete optimisation models of stop and station locations (see Furth et al., 2007; El-Geneidy et al., 2010).

The relaxation of the assumptions of parallel lines and spatially homogeneous demand leads to a substantially more complicated network optimisation task. The *transit route network design problem* (TRNDP) has a very populous literature in operations research. Due to the enormous size of the solution space when hundreds of nodes can be connected with each other in various combinations, this literature relies on heuristic algorithms, primarily. Reviews of the relevant literature are published by Kepaptsoglou and Karlaftis (2009) and Farahani et al. (2013).

Analytical investigations of optimal line structures have had to be limited to stylised networks. Jara-Díaz et al. (2012) considered simply two origin-destination pairs with an overlapping line section that can be served either with exclusive lines running in parallel or by introducing a trunk line with larger vehicles, enforcing passengers to transfer. Gschwender et al. (2016) extended this approach into a Y-shaped network where the trunk line serves three OD-pairs. Jara-Díaz and Gschwender (2003a,b), Jara-Díaz et al. (2014) as well as Jara-Díaz et al. (2018) analyse a network of one central and several peripheral nodes, with or without a financial constraint imposed on the operator. Fielbaum et al. (2017) introduce a parametric city model in which typical urban spatial structures can be reproduced by varying a small number of parameters. In this model, Fielbaum et al. (2016) derived the conditions under which either direct, feeder-trunk, hub and spoke and exclusive line structures minimise social costs. Fielbaum et al. (2018) established links with the TRNDP literature by implementing frequently used design heuristics in the parametric city concept. Fielbaum et al. (2020) studied the impact on (ray) scale economies of switching from one network layout to another as demand grows, showing that user benefits can be achieved as trip *directness* improves with fewer transfers in high-demand regimes. This extensive literature shares the same methodological foundations: demand is exogenous (parametrically defined), optimal frequencies and vehicle sizes minimise the sum of user and operator costs, urban spatial structure does not react to changes in transport supply, and mode choice is neglected. Daganzo (2010) took a different approach by considering a multimodal grid network in which he benchmarks optimally designed bus, BRT and metro technologies. Badia et al. (2014) adopted this analysis for the case of radial street networks.

3.3. Welfare-oriented pricing

Setting fares in public transport is a complex managerial and political process, where the range of conflicting objectives is probably even wider than in capacity optimisation. Pricing reforms may happen due to various financial, redistributional, and ideological reasons, and this review does not have the intention to cover all these motivations. Following the transport economics literature, we limit our attention to the welfare economics of pricing in public transport. In this context, the purpose of the fare is to manage demand such that only trips with a positive net contribution to social welfare get realised. From a more practical point of view, optimal fares ensure that (i) resources are efficiently utilised when public transport capacity is scarce, (ii) the external costs and benefits of travelling are internalised in the user's decisions, and (iii) sufficient revenues are generated if public funding has non-zero marginal cost for society. These goals are achieved in certain cases by simply setting the fare equal to the corresponding trip's marginal social cost. This is what the literature calls *first-best pricing* (Quinet, 2005; Small and Verhoef, 2007) and what Section 3.3.1 covers. If technological, institutional or financial constraints prevent marginal social cost pricing, then a constrained optimisation must be performed to derive the *second-best* fares (see Section 3.3.2). Finally, Section 3.3.3 reviews the literature of more advanced pricing schemes that diverge from linear usage-dependent fares.

3.3.1. First-best pricing rules

Further distinction is worth to be made between static and dynamic (time-dependent) pricing models. We begin our discussion with the static approach, where users' scheduling costs are neglected, and the objective is to maximise a time-invariant social welfare function. The principles of marginal cost pricing for congestible road use have a long history (see Lindsey, 2006), where it is well established that optimal tolls should reflect the marginal external congestion cost of driving. In the study of public transport pricing, due to the diversity in user and operator cost components, the social cost of the marginal trip may include several elements. The first pricing models in this mode emerge in the 1970s, with the works of Mohring (1972), Turvey and Mohring (1975), and Jansson (1979).

In the marginal social cost pricing paradigm, the structure and composition of pricing rules depends on what demand dependent costs are included in the welfare function, beyond the marginal traveller's personal cost. Certain marginal costs depend on ridership directly, while others accrue due to the responsive nature of capacity. Interestingly, the seminal model of Mohring (1972) belongs to the second group. When capacity is responsive, so that an increase in demand triggers capacity adjustment and thus a change in the cost borne by all other users, then it becomes clear that the operator's capacity policy will be a strong determinant of the optimal pricing rule as well. Mohring (1972) showed that under the optimal frequency rule introduced in Section 3.2.1, the operator cost of frequency adjustment induced by the marginal trip is equal in magnitude to the corresponding waiting time benefit enjoyed by all users. Such *indirect* benefits with responsive capacity normally point towards lower fares and lower self-financing ratios,[26] as Section 3.4 discusses in more detail.

The marginal trip may generate direct user externalities as well. The first example comes from Jansson (1980), who treated cycle time as a ridership dependent variable due to boarding and alighting time requirements. In this setup it is clear that the marginal user imposes an in-vehicle travel time externality on others. Prior to capacity adjustment, the marginal trip contributes to crowding, imposing an external crowding cost in fellow travellers. This setting resembles the textbook examples of static road congestion pricing. With responsive vehicle size and frequency adjustment, the externality can be internalised, transforming the inconvenience into operational expenses, but the net social cost of the marginal trip remains equal in magnitude to the marginal external crowding cost (Hörcher and Graham, 2018). Capacity adjustment might have a series of further social cost implications beyond the operator's expenditure. For example, high-frequency bus services that may slow down both cars and buses, as formally modelled by Else (1985), Basso and Silva (2014), and Tirachini et al. (2014a), and the capacity of a railway line may also saturate at some point (Pels and Verhoef, 2007), to mention two examples. Such indirect effects normally do not appear in the explicit formulae of welfare maximising pricing rules, but they may have an indirect impact through limited capacity adjustment and higher user costs due to crowding and failed boarding.

---

[26] Note that the terminology of "marginal cost pricing" is not precise in the sense that the marginal trip may generate social benefits as well. The external waiting time benefit caused by the Mohring effect is a typical example. Other positive externalities include agglomeration benefits investigated by Hörcher et al. (2020b) in the public transport context. Thus, in fact, the first-best policy equates the fare with the value of the marginal trip's contribution to social welfare above the marginal user's personal cost.

The second group of first-best pricing models adopts a dynamic approach, where users' departure time decisions are endogenous. This approach requires assumptions on the pattern of desired arrival times and schedule delay valuations among passengers – assumptions well-known from the dynamic road pricing literature (Vickrey, 1969; Arnott et al., 1990). In the road context, the dynamic pricing problem arises when insufficient capacity forms a bottleneck for the traffic flow. When demand exceeds the bottleneck capacity, travellers either have to depart earlier or later than preferred, or spend time in the queue upstream to the bottleneck. The purpose of dynamic pricing is to substitute queueing costs with a recyclable time-dependent payment. This way the bottleneck capacity remains fully utilised during the peak period, and drivers set their departure times optimally without queueing, due to the monetary incentive. Li et al. (2020) presented a comprehensive review of the evolution of scientific investigations of the bottleneck problem.

In public transport, queueing (failed boarding) has somewhat lower relevance in comparison with private road use, but scheduling decisions might remain sub-optimal without pricing incentives, because passengers neglect the crowding externality they impose on other users when they set their departure time (de Palma et al., 2017). Thus, time-dependent monetary incentives are equally relevant to achieve the socially optimal trip scheduling pattern. Interestingly, peak-load pricing for public transport specifically has been proposed very early by Vickrey (1955, 1963). Kraus and Yoshida (2002) and Yoshida (2008) showed this formally with fixed vehicle capacity and queueing delay only. Huang et al. (2005) replaced the assumption on fixed capacity with crowding discomfort, and Tian et al. (2007) extended the analysis to a many-to-one network (line). Two complementary contributions, de Palma et al. (2015) and de Palma et al. (2017), took a major step in understanding peak-period public transport supply by jointly optimising the number and timing of train departures, the fare schedule, and the interior seat supply of trains with a detailed crowding cost function. Their investigation is coupled with a welfare analysis of time-of-day-varying pricing, showing that the welfare gain does not depend on aggregate ridership, but the benefits of capacity expansion are greater under this pricing regime.

Is time-dependent pricing an effective way to alter passengers' trip scheduling decisions? The availability of smart card data enables researchers to perform *ex-post* empirical analyses of previously implemented pricing policies. In a recent study, Anupriya et al. (2020b) investigated an Early Bird Discount policy introduced in Hong Kong, in which the service provider offered a 25% fare discount for those who arrive at inner city commuting destinations before 8.15 in the morning. They find that in the first two months of policy implementation, the arrival time of regular commuters who enjoyed the discount has decreased by only 25 s on average, while the discount generated significant induced demand in the peak shoulder. This implies that, at least in the short run, substitution between departure time periods is much weaker than the own price elasticity of demand, and the non-negligible discount made little effective contribution to peak spreading.[27]

3.3.2. *Second-best pricing problems*

First-best conditions are not met in reality due to technological constraints and various distortions in the surrounding economy. Pure marginal cost pricing may be constrained by

(i) the presence of an inefficiently priced substitute, in the form of an alternative mode;
(ii) financial constraints imposed on the operator, either as an explicit budget constraint or in the form of costly public funds;
(iii) joint costs in service provision between different network segments or time periods and limitations in capacity adjustment; and
(iv) restricted capacity adjustment.

Some of these constraints are inevitable in practical policy problems, and therefore the role of first-best pricing rules reviewed in the previous section is limited to providing intuitive insights and a benchmark for second-best or other alternative policies.

The most common second-best constraint analysed is the case of public transport competing with underpriced private road use. If cars are underpriced, there is an excess demand on top of the welfare maximising travel flow, therefore it might be welfare improving to reduce the public transport fare under its first-best level in order to induce mode shift. This second-best mechanism is recognised consistently in models that include substitution with car use (see column 4 of Table A.1 in the Appendix). After density economies in user and operator costs, this provides another major theoretical support for public subsidies in industry. The optimal subsidy increases with (1) the degree of substitution between the two modes[28] and (2) the deviation between the actual road toll and its optimal value. From a different point of view, the introduction of efficient congestion pricing for cars reduces the financial pressure of public transport because an optimal road charge decreases the optimal subsidy (Small, 2008). In certain applications, it is more meaningful to derive the optimal fare subsidy, that is, the gap between the first-best and second-best fares, directly (Jackson, 1975). Jackson's case for second best welfare maximisation through an increase in the quality of bus service, as opposed to a fare reduction, parallels the contributions of Mohring (1972) and Turvey and Mohring (1975), who identify a first best justification for bus subsidies due to the reduction in bus waiting times if frequency is optimally adjusted as demand grows, which is another way of speeding up bus travel.[29]

Second-best public transport pricing when cars are priced at *average* instead of marginal cost has been analysed in the early 1970's by Sherman (1971) for one time period and by Glaister (1974) for two periods. Glaister (1974) found that a second-best bus fare below marginal cost is warranted not only in the peak but also in the off-peak period even if the off-peak is assumed to be congestion-free. This is because a low off-peak bus fare can attract peak car users and peak bus users, which in turn decreases the peak bus fare and further attracts some car users to public transport. Parry and Small (2009) obtained large gains in social welfare from diverting car users into public transport in peak periods, whilst the first-best argument to subsidise fares due to the reduction of users costs is larger in the off-peak.

One may criticise bi-modal models that ignore other modes beside private car use and public transport. In many applications, non-motorised alternatives such as walking and cycling are equally relevant substitutes of public transport, particularly for short trips, and thus second-best fares may attract users who would walk or cycle otherwise (Kerin, 1992). Tirachini and Hensher (2012) overcame this limitation by analysing a three-mode problem with cars, public transport and a non-motorised mode (assumed to be uncongestible). The second-best public transport fare may indeed be overestimated or underestimated, depending on the elasticity of substitution between each pair of the three alternative modes.

Multimodal pricing problems are considered in a dynamic context as well. In his analysis of the bi-modal equilibrium between private and public transport, Tabuchi (1993) assumed a dynamic bottleneck that

---

[27] Note, however, that time-dependent pricing is justified under the assumption of inelastic trip scheduling as well, provided that the marginal social cost of travelling varies over time.

[28] Substitution is clearly a heavily context and model dependent parameter that can be captured by demand systems based on a multivariate inverse demand function or a discrete choice specification (see Section 2.1).

[29] Interestingly, Jackson (1975) proposed a method to determine the optimal subsidy for bus speed improvements, instead of covering fare reductions, with the result depending on how large the increase in operator cost is to achieve improvements in speed.

arises when the flow of cars exceeds the capacity of the road (Vickrey, 1969). As demand grows it is more attractive to have a rail based alternative competing with cars, due to economies of scale in the former and congestion externality in the latter mode. Tabuchi's two-mode model has been subsequently extended by a number of researchers, including the consideration that rail is priced at average cost instead of marginal cost (Danielis and Marcucci, 2002), becoming second-best pricing, and that modal choice is governed by a logit choice model (Huang, 2002).

Returning to item (ii) in our list above, a financial restriction in a welfare oriented objective forms another major group of second-best pricing problems. Let us distinguish two ways in which financial restrictions may enter supply optimisation. In the first case, the targeted financial outcome is enforced by an explicit budget constraint, requiring full cost recovery (zero-profit constraint) or satisfying a non-zero financial target. In a model of demand-dependent operator costs, a zero-profit constraint requires the fare to be set equal to the *average* cost of operations. This implies lower-than-optimal ridership under scale economies, because the sum of marginal user and operator costs is less than the sum of the average costs (see Jara-Díaz and Gschwender, 2005). Jara-Díaz and Gschwender (2009a,b) incorporated a budget restriction into the derivation of optimal frequency, and show analytically that the impact of this constraint is equivalent to a reduction of waiting and in-vehicle travel time valuations on the user cost side.

In the alternative modelling approach, the financial result enters the welfare function with a multiplier capturing the marginal social cost of raising public revenues (e.g. Proost and Van Dender, 2008; Basso and Silva, 2014; Hörcher et al., 2020b). The multiplier is often called the *marginal cost of public* funds[30] (MCPF, Kleven and Kreiner, 2006). The MCPF is one plus the *tax revenue premium*, which measures the dead-weight loss of distortions caused by taxation in the markets where government revenues are raised. Empirical results show that the MCPF varies on a wide range depending on each country's tax system.[31] In a public transport supply model, the MCPF as an input parameter might have a huge impact on optimal outcomes, including fares, frequencies and the degree of subsidisation (Hörcher et al., 2020b). This implies that general "rules of thumb" cannot be provided for crucial policy questions such as subsidisation without considering the country of application.

Item (iii) in the list of second-best scenarios refers to joint costs between spatially or temporally differentiated markets. This natural phenomenon is hardly avoidable when fares, frequencies, the vehicle size and other supply variables cannot be differentiated between consecutive sections of a public transport line, because demand conditions are likely to differ on these markets. Rietveld and van Woudenberg (2007) investigated in a numerical setup the relative welfare benefit of relaxing such restrictions along a simple line serving three origin-destination pairs on two sections. They find that frequency differentiation would provide the highest benefit. Hörcher and Graham (2018) highlighted the importance of the *magnitude* of demand imbalances between jointly served markets. They show that a concentration of demand leads to higher second-best vehicle size and lower frequency, as the relative importance of crowding discomfort increases compared to waiting time. From a pricing point of view, in the absence of demand interactions, the optimal fare equals the marginal external crowding cost in each market. That is, peak fares should be significantly higher than in calmer markets (Jansson et al., 2015).

Capacity variables such as frequency and vehicle size might be restricted by technological conditions. For example, the signalling system may impose an upper bound of train frequencies, the capacity of bus stops sets a constraint for bus frequencies, and the dimensions of urban buses is restricted by regulation. If, for instance, the frequency constraint is binding, the operator has to react with a second-best vehicle size to variations in demand (see Hörcher, 2018, Section 6.3), and *vice versa*. The marginal social cost of a trips is likely to increase in this scenario, together with the socially optimal fare.

Finally, from a behavioural point of view, empirical evidence show that the performance of pure time-varying fares can be exceeded with a combination of *sticks* and *carrots*, that is, negative and positive incentives (Whelan and Johnson, 2004; Peer et al., 2016). As an example of positive incentives, Yang and Tang (2018) proposed awarding regular peak travellers with occasional free trips in the peak shoulders, which achieves peak spreading in a revenue neutral way. Tang et al. (2020) showed that the combination of regular uniform fares with an optional fare-reward scheme is Pareto-improving.

3.3.3. *Non-uniform pricing*

Non-uniform or non-linear pricing refers to "structures which permit us to vary prices not only between markets, but also between consumers in the same market" (Brown and Sibley, 1986). The most usual form of nonlinear pricing is the practice of quantity discounts, that is, second-degree price discrimination between frequent and infrequent users within a market. The simplest example is the combination of purely usage dependent fares (often called as *single tickets*, *pay-as-you-go fares* or *one-way tickets*) and subscriptions (*travel passes* or *season tickets*). Other forms of quantity discounts include *block pricing* which enables a given number of trips for a discounted price, without temporal limitations, and *price capping*, an upper bound on usage dependent spending, normally within a day. This subject is particularly relevant for public transport policy, given that in many cities around the world passengers are offered a wide range of substitute tariff products. The possibility of non-linear pricing is an interesting policy question in itself: travel passes are particularly popular in dense urban areas of Europe and North America, while such quantity discounts are often not available in Latin American and in several Asian cities. In a complex tariff system, each item might induce a different incentive mechanism for passengers, thus affecting both demand patterns, the operator's financial performance, and the overall economic efficiency of service provision. Therefore, it is surprising that the majority of research in the transport economics literature tends to focus on optimal supply with linear usage dependent fares only, and nonlinear pricing is vastly under-represented compared to its policy relevance.

Non-linear pricing gains attention in general microeconomic theory in the 1970s due to key contributions by Oi (1971), Littlechild (1975), and Leland and Meyer (1976). Carbajo (1988) was the first to adopt these developments for the case of public transport pricing, building on the modelling framework of the above cited papers. He shows that the combination of travel passes and single tickets allows for more efficient revenue generation compared to profit maximising flat fares. The reason is that travel passes enable the supplier to raise fares for infrequent users without discouraging frequent travellers' demand, for whom the subscription is an attractive alternative (Brown and Sibley, 1986). Even though profit maximisation is rare in our industry, the conclusions above imply that financially restricted public operators might also improve efficiency with non-linear pricing.

After almost three decades of silence,[32] Carbajo's model was

---

[30] Analytical links do exist between the two modelling approaches: the marginal cost of public funds in the second approach is equivalent to the shadow price of financial expenditures in the first one, if the exogenous financial constraint is set optimally. This relation is illustrated numerically by Sun et al. (2016).

[31] Basso and Silva (2014) reviewed a number of empirical estimates in the literature. The tax revenue premium ranges between 6.6% for Belgium (Proost and Van Dender, 2008), 20% for the average African country (Auriol and Warlters, 2012), and 30% in Sweden where distortionary tax rates are generally very high (Sørensen, 2010; Börjesson et al., 2019).

[32] In this period, nonlinear pricing related publications focused on demand estimation, without efficiency considerations; see e.g. FitzRoy and Smith (1998, 1999), Matas (2004), García-Ferrer et al. (2006), and Gkritza et al. (2011).

revisited by Jara-Díaz et al. (2016), who raised attention to the income effect in the travel pass problem. If a monthly season ticket, for example, constitutes a considerable proportion of a household's income, then it may shift individual demand for trips in this period; an aspect that the Carbajo neglected. Jara-Díaz et al. (2016) derives welfare maximising single ticket and travel pass prices with parameters from Santiago (Chile) under a zero-profit constraint, and concludes that the lowest eight out of ten income groups of the city should rely on passes, instead of usage fees. Hörcher et al. (2018a) extended Carbajo's model in another direction. They revealed a less desirable property of travel passes in a model that includes crowding discomfort as well. Given that pass holders face zero fare on the margin, while their trips may generate crowding externalities, an overconsumption threat arises. In a numerical model they show that the operator's financial constraint must be close to profit maximisation to justify the introduction of this kind of nonlinear pricing from an efficiency point of view.,[33][34] Hörcher and Graham (2020b) investigated the efficiency of subscriptions in a bi-modal setting, with underpriced private car use and endogenous car ownership. Using a three-level nested discrete choice demand system, they find that the introduction of subscriptions can make infrequent public transport users more reliant on their private cars, as they experience increased crowding due to the increased demand by pass holders. This implies that subscriptions induce welfare losses in both public transport and private car usage.

To the best of our knowledge, block pricing and fare capping have not been subjects of model-based microeconomic analyses so far. A frequent argument supporting fare capping is that it eliminates the uncertainty of subscription purchases, as users do not have to estimate in advance how many trips they are willing to undertake during the subscription's time interval (Chalabianlou et al., 2015). With fare capping, payments are usage dependent under the price of the corresponding subscription, and free above that. Thus, intuitively, it is plausible that fare capping is more efficient than a subscription, but it is unlikely that the social cost of the marginal trip drops to zero after the daily or weekly fare cap is reached, and this casts doubts about its superiority over pure usage dependent pricing. As fare capping is a popular element of several emerging policy initiatives such as *Mobility as a Service* (see Section 4.2), the need for more rigorous microeconomic analyses in this area is hardly questionable.

3.4. *Cost recovery and subsidisation*

The level of public transport subsidisation, or, in other words, the degree of cost recovery with fare revenues is an often debated subject in transport policy. There is a natural desire to develop *subsidy rules* that are able to tell the optimal financial contribution from the public budget, given the local characteristics of the public transport operator. In principle, this is not a straightforward task, because the optimal subsidy is just a derivative of the optimal pricing and capacity setting decisions. The previous parts of this section have illustrated that pricing and capacity rules can be rather complex, especially under second-best conditions; thus, the derivation of practice-ready, explicit subsidy formulae is not necessarily possible in reality. Naturally, empirical recommendations for the optimal subsidy in specific applications have developed in parallel with pricing and capacity literature.

To study the fundamental mechanisms behind the optimal subsidy, the Cost Recovery Theorem of Mohring and Harwitz (1962) is a useful starting point. The Theorem expresses the optimal degree of self-financing from (1) the degree of homogeneity of the demand a capacity function of user costs, and (2) the elasticity of the cost of capacity provision, assuming that capacity is indivisible. This rule is popular in the road pricing literature, where standard conditions (the user cost function is homogeneous of degree zero and capacity costs feature constant returns to scale. de Palma and Lindsey, 2007 reviewed the conditions that may distort full cost recovery, highlighting some of the specific features of operational and usage costs in public transport, including the Mohring effect. Hörcher and Graham (2018) adopted the Cost Recovery Theorem for public transport supply by considering the extreme cases when either waiting time or crowding discomfort is dominant on the user cost side. As the waiting time cost function is homogeneous of degree -1, while crowding inconvenience depends on the ratio of demand and capacity, these extrema lead to a self-financing ratio of zero and one, respectively. The user cost function is no longer homogeneous if both components are non-negligible, so the Mohring-Harwitz formula does not hold in this case. Nevertheless, one may infer that the optimal self-financing ratio depends on the relative magnitude of waiting and crowding costs. Börjesson et al. (2019) generalised this finding for waiting time and schedule delay costs on one hand, and congestion and other negative externalities (traffic crashes, noise, local pollution, climate gas emissions) on the other hand. They illustrate the interplay between waiting and crowding costs and its consequence of optimal subsidies in the context of a comparison between public transport provision in small and big cities.

3.5. *Private operations and public service regulation*

The industrial organisation of urban public transport has been dominated by publicly owned monopolies in many countries around the world. Nevertheless, interactions between multiple competing service providers or between an operator and its regulator is an increasingly relevant subject. One reason for this is that traditional public monopolies often face accusations of low internal efficiency compared to private firms, and therefore deregulation (competitive service provision) or a closer quality regulation under public service contracts can be used as policy tools to incentivise incumbent monopolies. Economic theory, and especially the fields of industrial organisation (IO) and game theory, provide a powerful toolbox for modelling the behaviour of the parties involved, optimising regulatory decisions, and uncovering the impact of earlier decisions. Under this section we focus on studies with modelling contributions and review three related subjects that attracted increased attention in the literature. These are (i) deregulation of the British bus industry in the 1980s, (ii) attempts to liberalise the conventional and high-speed passenger rail markets in the European Union since the 1990s, and (iii) contract theory behind public services, primarily in urban bus service provision.

The British urban and regional bus industry was the first to be opened up for competition in post-war European public transport industry (Evans, 1990). This major institutional change is among the most intensively documented ones in the transport economics literature, both prior to and after the intervention. In a heated academic debate, Mackie (1983), Gwilliam et al. (1985b), Beesley and Glaister (1985b), Gwilliam et al. (1985a), and Beesley and Glaister (1985a) reacted for and against the ongoing Government policy of on-the-road competition in local bus transport. Gwilliam, Nash, and Mackie argued that this market might not be contestable due to thin demand on certain routes, the advantage of incumbents and potential cartels, and even if an oligopolistic rivalry emerges, it is unlikely that it leads to allocative efficiency that would require subsidies (Mohring, 1972; Else, 1985). They proposed that even if private operators are more innovative and efficient internally, the benefits can be captured in competitive tendering as well, without the disbenefits of on-the-road competition. Beesley and Glaister responded that in a moment when the industry faces a crisis due to the cut in government subsidies, only deregulation is quick and effective enough to avoid a rapid deterioration of service quality and patronage, and the

---

[33] This finding resonates with Wang et al. (2011) who reach similar conclusions in the context of nonlinear road pricing with congestion externalities.

[34] Hörcher et al. (2018a) found that alternative management objectives (reviewed in Section 3.1) might explain why travel passes are widely used in industry: Substantially higher ridership can be achieved with this form of nonlinear pricing under a fixed budget constraint.

institutional costs of tendering should not be underestimated.

After eight and ten years, Nash (1993) and Mackie et al. (1995), respectively, documented that proponents of deregulation were right in the sense that operating costs have fallen more sharply than the opponents expected. However, the local bus market turned out to be imperfectly contestable and mergers and acquisitions prevailed over direct competition. White (1990) estimates that the net welfare effect was positive in metropolitan areas and negative in rural parts of the country where cost savings were more limited. The counterfactual case of London, where bus services were tendered instead of completely deregulated (Kennedy, 1995), shows that cost reductions could be achieved without the loss of ridership that other parts of the UK had experienced.

The British bus experience had a substantial impact on how the academic community approached the liberalisation of passenger rail services in the EU as a whole. Full privatisation of incumbent state-owned railway undertakings in combination with open-access competition had weak support in this mode. Due to the natural monopoly effect of large fixed costs, the first step towards a liberalised rail market was to separate infrastructure provision from train operations. The second step differed between freight and passenger service provision. In freight, on-track competition became the predominant market structure, leading to significant internal efficiency improvements and rising market shares (Nash, 2011). In terms of passenger services, the state of European railway liberalisation is fragmented. The United Kingdom went on its own path with the privatisation of the state-owned operator and the introduction of competitive tendering on a geographical basis. In other countries, incumbent national railways are mostly still in monopoly; competition is in some cases introduced by tendering local and regional services (e.g., Germany, the Netherlands, Denmark and Sweden) or allowing open access competition on the most profitable main lines (e.g., high-speed rail in Italy or intercity services in Austria and the Czech Republic).

How significant is the role that economic theory plays in European railway policy making? The area where quantitative methods are extensively used to support policies is the estimation of cost functions and testing for returns to density, scale and scope (see Section 2.2.2) in railway operations. This enabled informed decision making in terms of the restructuring of integrated railways and their separation into train operators and infrastructure managers (Di Pietrantonio and Pelkmans, 2004). Oum and Yu (1994); Gathon and Pestieau (1995); Cantos et al. (1999), and Cowie (2002) are among the key empirical studies published prior to the implementation of restructuring, focusing on the efficiency impact of vertical separation, while Preston (1996) focuses on horizontal separation, i.e. the optimal size of rail operations. The literature documents ex-post empirical studies as well, primarily on the impact of vertical disintegration on operating costs and productivity (Growitsch and Wetzel, 2009; Cantos et al., 2010; Mizutani et al., 2015) or a change in horizontal market structure on fares (Fröidh and Byström, 2013; Bergantino et al., 2015; Beria et al., 2016; Vigren, 2017). Very few quantitative analyses go beyond cost modelling to derive more specifically how interactions between horizontally competing or vertically separated firms depend on regulatory decisions. Exceptions include Preston (2008) and Ruiz-Rúa and Palacín (2013), who modelled on-track competition primarily by fares, which are then extended with timetabling considerations by Broman and Eliasson (2019) in the spirit of Evans (1987). Their simulation techniques allow the analyst to identify both the conditions under which competition prevails and the efficiency effects it leads to.

Competition, and in general the institutional separation of transport operators and authorities opens up a range of research problems in terms of the optimal contractual relationship between parties involved. The optimal contract provides the right incentive for an operator whose objective may differ from that of the regulator. In addition, information asymmetries may exist between the *principle* and the *agent* as the former has incomplete knowledge of the operator's productivity and cost-reducing effort. The theory of incentives addresses the resulting *adverse selection* and *moral hazard* between the contracting parties (Laffont and Tirole, 1993; Iossa and Martimort, 2011). Adopting the theory of incentives to public transport, a series of structural econometric analyses test the presence of inefficiencies in bus operating contracts with data from French cities (Gagnepain et al., 2011). The industry is characterised by two types of policies: *cost-plus* contracts that allow the reimbursement of all allowed production expenses, and *fixed-cost* contracts that incentivise operators to cut costs and thus increase the margin between a fixed subsidy and the operating loss. Calibrating a theoretical principle-agent model, Gagnepain and Ivaldi (2002) quantified the operators' inefficiency, the effort of managers, the cost of public funds, as well as the welfare loss of actual contracts compared to first-best policies. Gagnepain et al. (2013) moved this line of research into a dynamic context by considering the frequency of renegotiation of contracts between agencies and bus operators. They find that longer contracts (increased commitment) lead to higher welfare, but operators gain much more than the rest of society. Gagnepain and Ivaldi (2017) uncovered additional reasons why authorities may select types of contracts that are disadvantageous from an aggregate efficiency point of view. Their empirical results hint that pressure from local interest groups (i.e., trade unions and the operator's shareholders) as well as the authority's political agenda have a significant impact on the choice of contract types.

The economic models reviewed above work with general production functions and thus introduce a number of simplifications of public transport technology. In fact, the contracts applied in practice cover a much wider range of quality variables (e.g., vehicle maintenance, information provision, incentives for reliable operation and the behaviour of drivers). These dimensions of public transport supply are difficult to dealt with in the framework of incentive theory. Gómez-Lobo and Briones (2014) identified two main strands in the practice oriented (mainly descriptive) literature. The first one discusses more general aspects of market structure, competitive tendering, contract types, and asset ownership. The second one focuses on specific details of contract design, payment mechanisms, and the general trade-off between performance incentives and risk protection. Comprehensive reviews of the field are available in Hensher (2007), Hensher and Stanley (2010), Ponti (2011), Selviaridis and Wynstra (2015), and Hensher (2017b), among others.

3.6. Links with the urban economy

The literature reviewed so far investigates optimal supply in a partial equilibrium framework in which interactions between the transport sector and other markets in the surrounding economy are neglected. Let us now turn to general equilibrium approaches. Along the boundary between urban economics and transport science we find a number of research problems related to the impact of transport services on urban spatial structure, the housing market in general, households' choice of location for various activities, taxation at different levels of government, labour markets and urban productivity. The majority of these studies limit the representation of transport services to one mode only, which is normally (congestible) private road use. There are, however, a growing number of papers that consider public transport as a relevant model component, at least in a stylised way. The role of public transport is usually of the uncongested substitute for car use, which requires public subsidies to exploit its full potential in congestion alleviation. Our review suggests, however, that novelties in partial equilibrium public transport analysis are gradually spilling over into urban economics as well, and the gap between these two branches of the literature narrows.

Interactions with labour markets. Parry and Bento (2001) considered the interactions between transport and labour markets; car use is congestible, while public transport serves as a mode with a fixed price for users that can be reduced by fare subsidies. They analyse revenue recycling from congestion pricing, and find that public transport subsidies incentivise labour supply and reduce the dead-weight loss of labour taxes, but this form of revenue recycling is less efficient than directly cutting the distortionary labour tax. Tikoudis et al. (2015) extended this model framework with endogenous residential relocation in a monocentric city, and allow for increasing returns in public transport provision with the introduction of fixed operator costs. They assume average cost pricing in the baseline equilibrium, which is inefficient indeed in the presence of scale economies. Complementing the original findings of Parry and Bento (2001), they showed that if the fixed cost is above a threshold level, then turning revenues from congestion pricing into fare subsidies for public transport can be more efficient than cutting pre-existing labour taxes. Van Dender (2003) considered two transport modes and commuting as well as non-commuting trips. He shows that if the transport sector has no control over labour taxes, then, ideally, transport prices should be differentiated between trip purposes to avoid substantial welfare losses due to distortionary labour taxation.

Agglomeration economies, that is, productivity benefits from the densification of economic activity, are another reason why incentives to commuting can be justified. If labour supply and commuting are complements, then a positive externality weakens the traditional Pigouvian argument for transport taxes that internalise negative congestion (Arnott, 2007) or environmental externalities (Verhoef and Nijkamp, 2003). Hörcher et al. (2020b) implemented this idea in a public transport model specifically, where both commuting fares and the frequency of services are endogenous. They confirm in a simulation experiment that agglomeration economies have a positive impact on the optimal subsidy.[35]

City size and urban sprawl. A series of studies, including LeRoy and Sonstelie (1983), Sasaki (1990), and Brueckner (2005), analyse transport subsidies, mode selection in public investments, mode choice in daily commuting, and their impact on city size and the intensity of urban sprawl. Brueckner (2005) showed that subsidies have two main effects on city size. First, as they directly reduce the cost of commuting, they can indeed accelerate urban sprawl. On the other hand, subsidies are usually financed by taxes that reduce households' disposable income and, consequently, their demand for space. Su and DeSalvo (2008) presented empirical evidence to validate previous findings in this literature, showing that the urban area contracts with public transport subsidies and expands with a subsidy on car use.[36]

Endogenous location decisions. The choice of residential and working location is a relevant determinant of urban form in an inter-regional context as well. The key question is whether earlier findings on the links between transport supply, labour taxes and other policy variables hold if households are mobile in terms of residential and/or working locations. This is particularly relevant if pre-existing labour taxes distort such choices. Wrede (2001) showed that if both decisions are endogenous, then, in line with Parry and Bento (2001), it is efficient to deduct commuting expenses from the labour tax. This result no longer holds if residential location is flexible, but work takes place at a fixed location. Borck and Wrede (2009) shifted the focus from labour taxes to agglomeration economies, still in a duocentric setup in which wages are higher in one of the cities. Commuting subsidies internalise the agglomeration externality and shift workers to more productive regions. Subsidies to short-distance commuting *within* each city also do have benefits by recovering the first-best spatial allocation of residence, which is otherwise biased towards a sub-optimally low population in the productive city. However, Borck and Wrede (2009) acknowledged that they "do not believe that differentiated commuting subsidies exist because they are efficient". They proposed that "possible explanations rest on politics". We return to the multi-regional problem in Section 3.7, reviewing papers on the political economy of transport subsidies.

Computable general equilibrium models. Public transport as a separate mode appears increasingly regularly in large-scale spatial computable general equilibrium (CGE) models in which land use, labour markets, the transport network and other sectors of the local economy are integrated. The spatial scope of such models can be (i) limited to a single region, (ii) multi-regional, or (iii) focused on an urban area (Bröcker and Mercenier, 2011). Naturally, public transport has higher relevance in the last group of applications, and therefore it appears more often as a separate mode. The large scale of spatial CGE models currently does not allow for a detailed modelling of public transport specific components such as waiting times, crowding, etc. However, a major advantage of the CGE approach is its microfoundations, which allow for a welfare evaluation of specific interventions, even in a spatially disaggregated manner, and thus produce valuable outcomes for policy appraisal (Robson et al., 2018). One example is the Regional Economy Land Use and Transportation model of Anas and Kim (1996) and Anas and Liu (2007) that received a public transport module in more recent applications (see Anas, 2013). Also in a spatial CGE framework, Tscharaktschiew and Hirte (2012) analysed the welfare and redistributional effect of subsidies to public transport versus car commuting.

3.7. Political economy

Beyond more traditional welfare analyses, a growing body of literature investigates the political processes shaping decisions in public transport policy. Studies in this field are dealing with (i) intergovernmental competition in vertical or horizontal terms, that is, the difference between centralised and decentralised supply, (ii) deviations from the socially optimal supply if certain groups of passengers are in a majority within a jurisdiction, and (iii) the impact of the political outcome on aggregate social welfare and various groups of voters, including redistributional effects. Heterogeneity between residents is normally recognised in terms of the place of residence and working location, income, and car ownership.

When interest groups in society are separated by political (municipal) boundaries, then intergovernmental competition becomes increasingly relevant. One policy question is whether transport services are provided centrally by a federal government, or by regional entities in a decentralised manner. Federal service provision is generally closer to what we consider *socially optimal*, but in many applications the residents of the federation may be interested in taking control decisions on a regional level, or delegate this task to other regions. In modelling terms, this implies that the objective of supply optimisation changes from aggregate social welfare to a subset of its original components, representing the narrowed interest of low-level governments. The resulting supply rules may differ in various ways from the first-best optimum. First, a revenue generating motivation may appear in pricing, if residents of other regions are also among the users of a service supplied by one region. This is documented in the literature as a *tax exporting* behaviour (Oates, 1972; De Borger and Proost, 2016), which may, in case of public transport services under scale economies, turn into *subsidy importing* (Hörcher et al., 2020a). Second, decentralised decision making

---

[35] At the same time, Hörcher et al. (2020b) urge that mode shift (i.e. shifting road users to public transport) has no net agglomeration effect if commuters' contribution to urban productivity is independent of their mode choice. Thus, in a multimodal setup, if public transport is a close substitute of exogenously priced road use, then the impact of the agglomeration elasticity on optimal supply is much weaker.

[36] The urban economics of public transport provision has close links with the literature of urban planning. Key contributions, such those by Cervero (1998, 2004), summarise the main messages of *transit-oriented urban development*.

implies that not the full range of externalities are considered in the objective function. In particular, externalities borne by residents of competing regions are not internalised by local suppliers. Such spillovers may include consumption externalities as well as costs borne by third-parties, such as local pollution, accident risk, or positive agglomeration benefits, etc.

The problems above are dealt with a number of early papers focusing mostly on a congestible transport mode without specific technological details. For example,

- De Borger et al. (2007) established the principles behind regions' interest in differentiating the taxes imposed on local and transit traffic along a congestible inter-regional (or inter-national) corridor, and the way how investment in capacity is affected by their ability to do so.
- Vandyck and Proost (2012) integrated regional transport decisions with local labour markets, and discuss that local governments may have an interest to attract labour force from neighbouring regions, especially under agglomeration economies. Equivalently, less productive regions' interest is to retain labour by increasing the price of commuting and under-investing in infrastructure. These strategic considerations imply departures from the optimal allocation under decentralised supply.
- De Borger and Proost (2016) advanced the literature by modelling the (majority) voting process on both regional and federal levels, assuming heterogeneity in demand for a congestible transport service both within and between the regions. With this realistic representation of the political process, the superiority of federal decision making is no longer guaranteed. In fact, if users are in majority in at least one of the regions and transit traffic is not substantial, then decentralised supply is more efficient than federal control.
- Proost and Sen (2006) considered competition between vertically separated governments in the context of setting transport taxes (cordon tolls and parking charges, in particular). They found that the outcome of non-cooperative games is not far from the social optimum as the two fees are substitutes and the objective of the higher-level government partly overlaps with the local government.

Given that the cost structure of public transport provision differs from what these papers consider by focusing on road congestion primarily, the applicability of their conclusions for the political economy of public transport is ambiguous, *a priori*.

Politics behind public transport policy has received more limited attention in the literature. Brueckner and Selod (2006) and Brueckner (2005) investigated transport system choice from a continuum of monetary price and travel time combinations in an urban economy assuming majority voting. They concluded that cities with a heterogeneous skill and income distribution diverge from social optimum towards less expensive but more time-consuming modes. This finding reverses with the possibility of subsidisation: Brueckner (2005) showed that if residents are unaware of the link between transport system choice and their tax burden, then the city moves towards more expensive but faster modes, and smaller than optimal city size, especially if the decision is controlled by the rich group of residents. Brueckner (2005) claimed that this can explain overinvestment in freeways at the expense of public transport in the U.S., and he explains the compactness of European cities with lower transport subsidies.

Borck and Wrede (2005, 2008) analysed support of two income groups for commuting subsidies. They have a monocentric city in which the two groups live either in the central district or in the suburbs, but there is not political boundary between these areas. Subsidies for commuting impact the housing rents (see Brueckner, 2005, reviewed earlier), and thus landowners' income. Borck and Wrede (2005) derived the impact of subsidies on group-level welfare in function of their equilibrium residential location and (exogenous) share among landowners. This paper neglects modal specificities in the model, however. Borck and Wrede (2008) took a step forward in this sense by introducing two transport modes. A general finding is that if land is owned by residents, then subsidies redistribute between the two income groups; otherwise, all residents gain from lower rents at the expense of landowners. Group-level preferences are derived for specific residential patterns: the paper shows that if the poor live in the centre and the rich commute from the suburbs, for example, then public transport subsidies hurt the rich, but supporting car commuting may benefit the poor through lower land rents in the city centre. De Borger and Proost (2015) returned to the two-region setup with two modes and two groups of voters: car owners and non-car owners. This is the first political economy model with a detailed specification of user costs in public transport, including the cost of waiting, delays due to boarding and alighting, and crowding on board. The authors focus on decentralised supply decisions in one of the regions, optimising public transport fares, frequencies and road tolls. The results are partly driven by tax exporting: The higher the share of foreign users, the higher the fare and toll levels set by the local provider. De Borger and Proost (2015) showed the general finding that cost recovery ratios and fares are higher under a decentralised government, and this tendency strengthens with the share of non-local users.

Even though most of the theoretical studies reviewed above are indeed motivated by the researchers' empirical observations in the policy sphere, robust statistical analyses with data from a larger pool of comparable areas of application rarely support their empirically testable hypotheses. This creates an obvious research gap for future empirical contributions.

3.8. *Social equity and distributional concerns*

Pricing decisions in public transport not only have welfare implications in absolute terms, but also distributional effects. Social equity in transport pricing is usually a salient concern of policy makers and elected public servants, with a larger relevance than the economic efficiency of a proposed reform in several cases (Quinet, 2005), to the point that distributional issues can be an argument not to implement marginal cost pricing in public transport. In spite of the relevance of social equity concerns, the distributional effects of alternative pricing decisions are often ignored in theoretical models of optimal pricing models, either due to methodological complexities, or because differences in distributional effects are considered not so relevant. The omission of an equity analysis might have adverse consequences for the practical applicability of research findings in public transport economics.

The main source of methodological complications is that full distributional analyses require (i) user heterogeneity in the model and (ii) a general equilibrium approach. Note that even if a model features multiple income groups, for example, there is no straightforward way in which operating costs can be assigned to individual users of jointly served public transport markets. There is a clear difference between the average and marginal operating cost of a trip due to scale economies and cost inter-dependencies (see Section 2.2.2 and Basso et al., 2011b). The choice between the two when disaggregate subsidies are calculated is arbitrary.[37] Second, a general equilibrium approach would be necessary to identify who is affected by, for example, an increase in the labour or land properties tax to finance and increase the subsidy for public transport (Dodgson and Topham, 1987; Proost et al., 2007). Dodgson and Topham (1987) found that the existence of social benefits of fares subsidies financed by additional taxes depends on the income elasticities of demand for private and public transport and for the taxed good.

The equity and distributional effects of public transport pricing and subsidy policies have been analysed in several partial equilibrium

---

[37] For example, if one income group has a disproportionate share in peak travelling, then the subsidy computed on an average cost basis underestimates the extent to which they are supported by society.

studies. Findings vary a great deal depending on methodology, the income profile of public transport users, and on which mode of transport is included in the analysis (bus and/or rail). Basso and Silva (2014) estimated that bus subsidies are progressive in Santiago, similar to Matas et al. (2020), who found that public transport subsidies (including bus and rail) are progressive in Barcelona. Börjesson et al. (2020) estimated that the average public transport subsidy per person is very similar for all income groups except for the highest income quantile, concluding that subsidies are not effective as a redistributional policy in Stockholm, where public transport is more uniformly used across income groups than in Barcelona and Santiago.

Many equity analyses are driven by the spatial distribution of low- and high-income households in cities. If lower income households are predominantly located in the outskirts, they experience the longest travel distances in public transport. This makes distance-based pricing rules unattractive from an equity point of view, despite their superiority in terms of economic efficiency. Due to differences in urban spatial structure, flat fares are shown as more equalitarian in Santiago (Tiznado-Aitken et al., 2020) and Stockholm (Rubensson et al., 2020), while distance-based fares benefit lower income people the most in Utah, according to Farber et al. (2014). Yet a different conclusion for Barcelona was reached by Matas et al. (2020), who found that flat and distance-based public transport pricing structures have rather homogeneous effects on social equity. Another relevant result regarding income profiles and mode choice is provided by Iseki and Taylor (2010), who estimated public transport subsidies to be regressive in Los Angeles, because low-income short-distance passengers that mostly use buses benefit less from the current subsidy policy than high-income long-distance commuters that mostly use rail. The geographical distribution of public transport users across a city, particularly regarding household income levels, plays a key role in these contrasting results.

The equity-oriented optimisation of supply policies is an emerging area in the literature. If public transport use is concentrated in low- and middle-income households, social equity concerns can significantly change the economic assessment of transport taxes and subsidies. Using Santiago data, the bimodal bus-car model of Tirachini and Proost (2021) shows that a welfare improving reform would increase the car cost and reduce the bus fare in peak periods, and would reduce the car cost and increase the bus fare in off-peak periods. However, the existence of distributional concerns operationalised through an income inequality aversion parameter (following Mayeres, 2001) implies that having lower bus fares and higher car cost in both peak and off-peak periods increases social welfare.

All in all, no rule can be considered as generally applicable to all cities: the equity effect of a given public transport pricing structure is highly dependent on the local social context and land use policy. Redistributional results averaged across income groups or geographical locations can hide important insights (e.g., low-income households living in the city centre receive a different level of public transport subsidy from those living in the outskirts). Such local discrepancies support the use of fine-grained spatial models for the social equity analysis of public transport pricing and subsidies, which are becoming increasingly more used due to the greater availability of big data on trip-level public transport use.

4. Emerging technologies and related policies

We devote a separate section in this review to provide an overview of new technological solutions and their economic properties. New technologies affect the operational process of public transport provision as well as the user experience and demand levels. The first subsection of this subject deals with innovative ways in which traditionally fixed schedules and routes can be made more flexible and adaptive to external conditions. Section 4.2 discusses new contributions at the boundary between public transport and emerging modes of the sharing economy paradigm. Section 4.3 reviews the most recent literature of automated public transport.

4.1. *Demand-responsive public transport*

In order to exploit economies of scale in high-capacity vehicles, traditional public transport technology is based on fixed routes, stations/stops, and predetermined timetables or service frequencies. We saw earlier in Section 3 that in the presence of spatially and temporally fluctuating demand conditions, such technological constraints imply efficiency losses, meaning that service frequencies and vehicle capacities can never match the their first-best optimal in all markets served. It is a natural desire to relax some of the constrained technological features of public transport, thus making it more adaptive to changing demand conditions.

Flexible transport services (FTS) or demand-responsive transport (DRT) allow for flexibility in either the routing of shared vehicles, or the timing of departures including the regularity of service provision, or the locations where passengers can board and alight, or the size of vehicles, or any combinations of these flexible features. As we reduce the number of predetermined supply variables, the boundaries between public transport and what we call taxi, ride-hailing or shared ride-hailing becomes more obscure. The generic trade-off as we move along this scale is between the returns to scale in the number of travellers sharing the same vehicle at the same time, and the convenience and speed of displacement when the shared vehicle's trajectory does not need to be adjusted to the personal needs of many users. Thus, the optimal balance between vehicle size and flexibility requires important policy decisions (addressed by Li and Quadrifoglio, 2010 and Navidi et al., 2018, among others), where the spatial density of demand has a key role (Quadrifoglio and Li, 2009) as well as observed traffic and congestion levels. Some economic characteristics remain persistent in the entire range between public transport and taxis, however. One of them is the presence of economies in the spatial density of the shared vehicles of a given mode, affecting the user cost of accessing the system (i.e., walking and waiting times), which in most cases is directly related to the intensity of use (see the links between Mohring, 1972 and Arnott, 1996).

The dial-a-ride problem and its evolution. The optimisation of flexible transport supply is a challenging task from a methodological point of view. The *dial-a-ride problem* (DARP) in operations research covers various optimisation techniques for the derivation of supply in a general FTS, focusing primarily on capacity, route planning and timetabling. The aim in DARP is "to plan a set of minimum cost vehicle routes" capable of accommodating trip requests between a set of origins and destinations, under a set of additional constraints, e.g. in terms of trip timing. Cordeau and Laporte (2007), Molenbruch et al. (2017), and Ho et al. (2018) provided in-depth reviews of the field. Pricing of DRT services is much less intensively discussed in the literature, possibly due to its methodological challenges: in a spatially disaggregated demand system, straightforward pricing rules cannot be derived analytically. In addition, in most of the DRT applications demand density is very low, demand on a given origin-destination pair might have to be reduced to the binary decision of a single potential user. The above cited operations research literature normally considers exogenous demand and a cost minimising objective. There are exceptions, however: Amirgholy and Gonzales (2016) presented a queueing model of dynamic DRT demand in which pricing tools are used to regulate the temporal distribution of requests.

Frequent applications of the dial-a-ride problem include special transport services (STS) for the elderly or disabled (Brake et al., 2007; Nelson et al., 2010), paratransit, internal transport services within airports (Reinhardt et al., 2013) or hospitals (Beaudry et al., 2010), and public transport in low demand periods (Mulley and Nelson, 2009), notably night services (Parragh et al., 2015) or rural mobility (e.g., Garaix et al., 2011).

Developments in information and communication technologies (ICT) open up the possibilities of smaller adjustments of traditional scheduled public transport, depending on demand variations. *Semi-flexible* services or *demand adaptive systems* (DAS) feature demand dependent adjustment capabilities in one or several planning dimensions (Malucelli et al., 1999; Nourbakhsh and Ouyang, 2012). Koffman (2004) and Potts et al. (2010) combined line segments with predetermined route and stop locations with another part of the line remaining flexible in these dimensions. This approach can be generalised by allowing *any* part of the line to be adjusted to demand shocks. Inspired by Koffman (2004) and Potts et al. (2010), Errico et al. (2013) presented a unifying model framework for semi-flexible services, which combines the advantages of regularity in scheduled public transport with the possibility of more personalised service provision, at the cost of additional stops or detours along a partially pre-planned line. However, little is known about the economic consequences of adding flexibility to traditional public transport, as the literature reviewed by Errico et al. (2013) focused on solution methods for capacity optimisation and scheduling.

Integrating scheduled and demand-responsive services. Integration between scheduled, fix-route public transport and demand responsive feeder services is also gaining increased attention in the literature. In the *Integrated-DAR* (Integrated dial-a-ride, or IDAR) version of the dial-a-ride problem many users share the same trip origin or destination (the closest public transport station), and trip timing is also important to offer smooth transfers to scheduled services (Kim and Schonfeld, 2014). This implies that demand is clustered temporally as well, thus improving the efficiency of vehicle sharing. Often cited methodological contributions in the literature are Hickman and Blume (2001) and Häll et al. (2009).

4.2. Sharing economy and mobility as a service

Over the past decade, technological development, particularly the mass adoption of low-cost information and communication technologies, has unlocked a series of new ways of sharing small-sized urban transport vehicles. Sharing offers the convenience of individual travel without the need of vehicle ownership, and thus a greater flexibility in the use of cars or bicycles, for example, in daily trip chains. Car sharing and ride hailing became a substitute of private car use and taxi travel, primarily due to the capital and labour cost savings they provide, and the efficiency gains provided by real-time driver-customer matching algorithms. From the viewpoint of public transport, the success of sharing economy raises two highly relevant policy questions: (i) whether the new modes are substitutes or complements of traditional public transport, and (ii) whether the public transport industry could adopt some of the emerging technologies to improve its own efficiency, and thus strengthen complementary demand effects against substitution.

Recent contributions in the literature address question (i) with empirical tools. They all agree in that the impact on public transport is very context dependent; in areas with poor accessibility, substitution tends to be stronger, while frequent public transport services may benefit from new solutions that play a feeder role. A synthesis of these empirical studies is provided by Tirachini (2020). Hall et al. (2018) estimated the causal effect of ride hailing on public transport ridership in a difference-in-differences setup, using the fact the major service provider entered the market in various cities gradually, and with different intensity. They find that in the average US city, ride hailing increased ridership by 5%, and due to heterogeneity in causal effects, smaller cities with relatively larger transit agencies might have experienced degrading patronage.

The success of sharing economy (and partly its competitive threat to existing modes) has inspired a new wave of technological innovations and policy initiatives. *Mobility as a Service* (MaaS) is one of the most influential concepts aiming at an improved integration of all shared and public transport modes. MaaS has no consensual definition in the literature; it has to be defined as a collection of transport policies instead of an individual mode of transport. MaaS policies promote multimodality with a mix of traditional transport modes and sharing economy, building heavily on online communication technologies in travel information provision as well as pricing. The concept has the potential of becoming an alternative for individual car ownership. Early publications, including Aapaoja et al. (2017), Hensher (2017a), Smith et al. (2018) and Sochor et al. (2018), cite partial empirical evidence in support of this proposition, showing that MaaS products are perceived positively by car owners in stated preference experiments. The MaaS concept has a strong institutional dimension as well. It proposes that a central agency, the *MaaS provider* should play the coordinator's role by collecting the service offers of the operators of various modes and presenting them as a unified menu of service for passengers.[38]

Another innovative feature of MaaS is the integration of the tariff systems of incumbent service providers into *bundles* or *mobility packages* of travel permits, reformulated as multimodal subscriptions (Matyas and Kamargianni, 2018; Mulley et al., 2018; Guidon et al., 2020). The impact of MaaS tariff products on travel demand is estimated in various choice experiments by Wong et al. (2018); Matyas and Kamargianni (2018); Caiati et al. (2020); Guidon et al. (2020), and Ho et al. (2020). Hörcher and Graham (2020b) showed that even though multimodal passes are effective in price discrimination and revenue generation, their welfare effect might be detrimental in the presence of crowding and congestion externalities.[39] They propose that disaggregate (spatially and temporally differentiated) pricing serve the general societal goals of MaaS more effectively. We observe growing interest around MaaS among industry stakeholders as well as researchers. However, most of the MaaS proposals are not supported with solid theoretical and empirical evidence, and therefore we see substantial room for future contributions in this field.

4.3. Automated public transport

The automation of road vehicles is expected to have profound impact on mobility, including the future of public transport. Driverless trains are already widely adopted in urban metros and suburban train services, where the physical constraint in the rail-wheel contact and the segregation of railway infrastructure make automation a feasible task at the current phase of technological development. Automation enables even more promising prospects for bus services, as in this mode, the driver's wage constitutes a very significant fraction of the total operating cost (larger than in the case of rail services), in the range between 40% and 70% in developed economies such as Sweden, Australia and Singapore (Jansson, 1980; Ongel et al., 2019; Tirachini and Antoniou, 2020) and around 30% in developing countries such as Chile (Tirachini and Antoniou, 2020). Even though little is known about the capital cost of automated bus operations when automation technology becomes mature enough, and the reaction of demand is also unknown beside the lessons learnt from small-scale pilot projects (see, e.g., Ainsalu et al., 2018), some early contributions already do provide guidance in terms of the potential economic impacts of automated public transport.

Wadud (2017), Bösch et al. (2018) and Ongel et al. (2019) estimated

---

[38] Several elements of the MaaS concept reflect earlier propositions of transport policy. For example, Nelson et al. (2010) proposed a *Flexible Agency for Collective Mobility Services* as "an organisation structure and business model" for flexible transport services, several years before the debut of the MaaS concept.

[39] See a more detailed discussion on non-linear pricing aspects in Section 3.3.3 of the present paper.

the operator cost savings that taxi and bus automation would imply and showed that the cost reduction might be very substantial. Wadud (2017) finds that 60 percent of bus drivers' current wage expenditure could be saved in the UK; Bösch et al. (2018) claimed that the cost of taxi trips could drop by not less than 85% in Zurich, while Ongel et al. (2019) estimated a 70% reduction in the cost of operating 6 m long shuttle buses in Singapore. Four recent studies go beyond the operator's viewpoint, analysing how automation affects user well-being. Abe (2019) identified the impacts on the in-vehicle travel time of passengers of automated taxis and buses in Japan. Tirachini and Antoniou (2020) showed that within a frontier of driver cost reduction and running speed limitation, the optimal size of automated buses might be smaller than traditional ones, allowing for higher frequency and less waiting time on average. They show that after the disappearance of the fixed cost of drivers, the degree of density economies in bus operations becomes milder, and therefore cost recovery should improve under optimal pricing, thus reducing the need for public subsidies. Zhang et al. (2019) found that semi-automated buses capable of automatic platooning can deliver partial savings, especially on interurban lines where the distance between stops is relatively high and demand is sparser. Finally, Fielbaum (2019) put this new technology into a network context, and shows using a general urban spatial structure that automation might lead to more direct routes compared to the feeder-trunk system that characterises many current bus operations.

The story of automated public transport is not expected to end here. Key questions are (i) how congestion technology will evolve with automated vehicles, (ii) how automated vehicles will interact with other travellers including pedestrians and cyclists, and (iii) if passengers are going to accept boarding driverless buses. Without substantial improvements in car following capabilities, the prediction of higher frequency and smaller bus size is hardly plausible in dense urban areas. On the other hand, automated control might enable operators for example to improve headway regularity more effectively, thus eliminating the adverse effects of bus bunching at high frequency. The spread of flexible, semi-flexible and demand adaptive operating strategies, reviewed in Section 4.1, may also be facilitated by automation and the fact that entire fleets could be controlled by one routing and scheduling algorithm. As soon as the technological details of feasible mass-produced automating solutions unfold in the future, a lot more will be learnt about their economic properties as well.

5. Conclusions

This paper reviews more than 300 of the most influential studies on public transport economics. In terms of the methodologies they were built on, we conclude that our field follows and oftentimes even forms the state-of-the-art of the general economics literature. The list of the most frequently investigated subjects is dominated by the optimisation of pricing and subsidy policies, where scale economies in user and operator costs and substitution with underpriced private car use are the leading mechanisms behind incomplete self-financing in social optimum. We see a number of directions in which the literature shows ongoing development and where more research would be needed to achieve stronger impact in policy making.

We believe that the discussions and debates on optimal subsidies are by far not over. In the late 20th century this subject was relevant because, at the pinnacle of private car use, public transport needed theoretical support to justify that some level of *subsidisation* is not distortionary in a competitive intermodal transport market. Today we observe the opposite tendency as well: Economists also have to justify that some sort of *pricing* is still needed to ensure the efficient use of resources allocated to public transport provision even if this mode is perceived as more sustainable than private car use. The ideologies of "zero subsidy" and "zero fare" are both present in today's global transport policy, with their relative power varying widely between continents and countries. Relatively new and sometimes off-setting factors of the optimal subsidy now include crowding externalities, network-level spillovers, the marginal cost of public funds and potential agglomeration economies.

Within the sphere of the studies that investigate public transport in isolation, we see a tendency that models of the representative OD pair are getting replaced with line- or even network-level representations in which operational and demand characteristics can be treated in a disaggregate manner. The dilemma of the balance between analytical transparency and practical applicability will remain with us in the future, but it is important to leverage on the availability of computing power and adopt theoretical models in large-scale quantitative simulators as well.

Modelling interactions with substitute and complementary modes will remain high on the research agenda. The literature features a range of microeconomic methods to replicate modal substitution (see our summaries in Section 2.1 and Table A.1). We see room for unifying research efforts to better understand how these demand systems perform in reproducing what we observe in reality. Quantitative studies rarely perform sensitivity checks with respect to model selection, and as simulations diverge from currently observed demand levels and elasticities, the reliability of model predictions fade considerably. In terms of applications, substitution between car use and public transport was the top hit in the literature in recent decades. However, this focus may shift in the future to interactions with shared modes and personal mobility, against which the superiority of public transport in terms of negative externalities is no longer trivial, and substitution effects may also be stronger.

At the same time, one cannot be fully satisfied with the impact that the literature of public transport economics has on practical decision making. Theoretically, this mode is easy to regulate, as the sector is mostly characterised by public monopolies. Still, we rarely see that decisions are made according to the abstract benchmark of the maximisation of economic welfare. A better understanding of political processes behind transport policies is inevitable to bringing research findings closer to social acceptance and actual implementation.

This paper is completed in the middle of the COVID-19 pandemic, when the global public transport industry is experiencing an unprecedented shock, tremendous loss in demand, and severe financial difficulties.[40] Some commentators express doubts about the future of public transport. Our position is that public transport cannot be replaced by alternative motorised transport modes — especially not in densely populated urban areas — due to the *space-hungry* nature of all available alternatives, and especially not when the public health crisis leaves a large fraction of society in poverty. However, if the threat of infectious diseases becomes permanent in a globalised world, our prognosis is that demand management and efficient resource allocation, which are the key subjects of public transport economics, will become more relevant than ever.

CRediT authorship contribution statement

Daniel Hörcher: Conceptualization, Methodology, Writing – Original draft, review, and editing, Visualization. **Alejandro Tirachini:** Conceptualization, Methodology, Writing – Original draft, review, and editing, Funding acquisition.

Acknowledgement

This research has been partially supported by ANID Chile (Grant PIA/BASAL AFB180003). We are also grateful to two referees and an editor of this journal for many constructive remarks that much improved the final version of this paper. We take responsibility for any remaining shortcomings.

---

[40] See recent empirical evidence on the impact of COVID-19 on public transport demand and operations in the early assessment of Tirachini and Cats (2020), the disaggregate demand data of metro operators by the Transport Strategy Centre (2020), and a review of public transport specific epidemiological findings by Hörcher et al. (2020c).

Appendix

A Literature Overview

Table A.1
Travel modes and the demand system

	Modal Setup				Demand Systems								
	Bus	Rail	Other/ No tech details	Substitution with private mode	Parametric demand	Inverse/ direct demand function	Benefit function of trip volumes	Discrete choice model	Heterogeneous individual demand	Endogenous departure time	Utility from leisure & consume	Endogenous location choices	Wider economic effects
Mohring (1972)	✓				✓								
Glaister (1974)	✓			✓			✓						
Jansson (1980)	✓				✓								
Kocur and Hendrickson (1982)	✓			✓		✓		✓					
Carbajo (1988)			✓			✓			✓				
Oldfield and Bly (1988)	✓					✓							
Chang and Schonfeld (1991)	✓				✓	✓							
Kraus (1991)		✓			✓								
Jansson (1993)	✓					✓			✓				
De Borger et al. (1996)			✓	✓			✓						
De Borger and Wouters (1998)	✓			✓			✓						
Gronau (2000)	✓			✓	✓				✓				
Huang (2000)		✓		✓	✓				✓	✓			
Parry and Bento (2001)		✓		✓							✓		✓
Kraus and Yoshida (2002)		✓			✓	✓				✓			
Rietveld et al. (2002)		✓			✓	✓							
Jara-Díaz and Gschwender (2003a,b)	✓				✓								
Small (2004)	✓			✓	✓	✓							
Pels and Verhoef (2007)		✓		✓			✓						
Rietveld and van Woudenberg (2007)		✓			✓	✓							
Proost and Van Dender (2008)	✓			✓			✓						
Ahn (2009)	✓			✓			✓						

(*continued on next page*)

Table A.1 (*continued*)

	Modal Setup				Demand Systems								
	Bus	Rail	Other/ No tech details	Substitution with private mode	Parametric demand	Inverse/ direct demand function	Benefit function of trip volumes	Discrete choice model	Heterogeous individual demand	Endogenous departure time	Utility from leisure & consume	Endogenous location choices	Wider economic effects
Jara-Díaz and Gschwender (2009a,b)	✓				✓								
Parry and Small (2009)	✓	✓		✓			✓						
Calthrop et al. (2010)		✓		✓							✓		✓
Basso et al. (2011a)	✓		✓	✓				✓	✓				
Basso and Jara-Díaz (2012)	✓			✓	✓			✓	✓				
Basso and Silva (2014)	✓			✓				✓	✓				
Tirachini et al. (2014b)	✓			✓				✓	✓				
Tirachini (2014)	✓				✓								
De Borger and Proost (2015)			✓	✓			✓		✓			✓	
de Palma et al. (2015)		✓			✓			✓	✓	✓			
Fielbaum et al. (2016)	✓				✓								
Gschwender et al. (2016)	✓				✓								
Börjesson et al. (2017)	✓			✓			✓						
de Palma et al. (2017)		✓		✓	✓	✓				✓			
Hörcher and Graham (2018)		✓			✓	✓							
Hörcher et al. (2020b)		✓		✓							✓		✓

Table A.2
User costs, temporal and spatial differentiation of service provision, including network layouts

	User Costs									Time		Spatial Layout						
	Access & egress walk time	Waiting time	Schedule delay	In-vehicle travel time	In-vehicle crowding	Station crowding	Denied boarding	Information collection	Transfer penalty	Peak time & off-peak time	Multiperiod services	Representative (aggregate) OD	Demand imbalance factor	Backhaul problem	Line with multiple sections	Urban space with uniform demand	Monocentric city corridor	Network with transfers
Mohring (1972)	✓	✓								✓		✓		✓				
Glaister (1974)				✓						✓		✓						
Jansson (1980)		✓		✓						✓		✓						
Kocur and Hendrickson (1982)	✓	✓		✓												✓		
Carbajo (1988)										✓		✓						
Oldfield and Bly (1988)		✓		✓			✓						✓					
Chang and Schonfeld (1991)	✓	✓		✓							✓					✓		
Kraus (1991)				✓	✓												✓	
Jansson (1993)			✓	✓	✓			✓		✓		✓						
De Borger et al. (1996)				✓						✓		✓						
De Borger and Wouters (1998)		✓		✓						✓		✓						
Gronau (2000)		✓		✓								✓						
Huang (2000)		✓	✓		✓	✓					✓	✓						
Parry and Bento (2001)				✓								✓						
Kraus and Yoshida (2002)			✓				✓					✓						
Rietveld et al. (2002)		✓		✓								✓						
Jara-Díaz and Gschwender (2003a,b)		✓		✓	✓							✓						
Small (2004)	✓	✓		✓												✓		
Pels and Verhoef (2007)		✓		✓	✓						✓				✓			
Rietveld and van Woudenberg (2007)		✓		✓											✓			
Proost and Van Dender (2008)		✓		✓						✓		✓						
Ahn (2009)		✓		✓								✓						
Jara-Díaz and Gschwender (2009a,b)		✓		✓								✓						
Parry and Small (2009)	✓	✓		✓	✓					✓		✓						
												✓						

(*continued on next page*)

Table A.2 (*continued*)

	User Costs									Time		Spatial Layout						
	Access & egress walk time	Waiting time	Schedule delay	In-vehicle travel time	In-vehicle crowding	Station crowding	Denied boarding	Information collection	Transfer penalty	Peak time & off-peak time	Multiperiod services	Representative (aggregate) OD	Demand imbalance factor	Backhaul problem	Line with multiple sections	Urban space with uniform demand	Monocentric city corridor	Network with transfers
Calthrop et al. (2010)																		
Basso et al. (2011a)	✓	✓		✓								✓						
Basso and Jara-Díaz (2012)		✓		✓								✓						
Basso and Silva (2014)	✓	✓		✓						✓		✓						
Tirachini et al. (2014b)															✓			
Tirachini (2014)	✓	✓		✓							✓	✓			✓			
De Borger and Proost (2015)		✓		✓	✓													
de Palma et al. (2015)			✓		✓							✓			✓			
Fielbaum et al. (2016)		✓		✓					✓									✓
Gschwender et al. (2016)		✓		✓					✓									✓
Börjesson et al. (2017)	✓	✓		✓	✓				✓		✓							
de Palma et al. (2017)			✓		✓							✓						
Hörcher and Graham (2018)		✓		✓	✓						✓			✓				
Hörcher et al. (2020b)		✓		✓	✓							✓						

Table A.3
Transport technology

	Operator cost function								Tech. details			
	Demand dependent	Frequency dependent	Vehicle size dependent	Fleet size dependent	Just infrastructure cost	Temporal cost interactions	Spatial cost interactions	Costly public funds (MCPF)	Endogenous cycle time	Interaction with congestion	Vehicle capacity constraint	Bus stop congestion
Mohring (1972)		✓		✓					✓			
Glaister (1974)	✓					✓				✓		
Jansson (1980)	✓	✓	✓	✓		✓			✓		✓	
Kocur and Hendrickson (1982)				✓							✓	
Carbajo (1988)	✓											
Oldfield and Bly (1988)		✓	✓						✓	✓	✓	
Chang and Schonfeld (1991)		✓		✓					✓		✓	
Kraus (1991)	✓								✓			
Jansson (1993)	✓	✓							✓			
De Borger et al. (1996)	✓							✓		✓		
De Borger and Wouters (1998)	✓	✓		✓		✓		✓	✓	✓		
Gronau (2000)		✓	✓						✓	✓		
Huang (2000)*												
Parry and Bento (2001)*												
Kraus and Yoshida (2002)		✓	✓	✓							✓	
Rietveld et al. (2002)	✓	✓	✓								✓	
Jara-Díaz and Gschwender (2003a, b)		✓	✓						✓			
Small (2004)		✓	✓					✓		✓	✓	
Pels and Verhoef (2007)	✓	✓	✓				✓			✓		
Rietveld and van Woudenberg (2007)	✓	✓	✓								✓	
Proost and Van Dender (2008)								✓		✓		
Ahn (2009)		✓								✓		
Jara-Díaz and Gschwender (2009a, b)	✓	✓	✓					✓	✓		✓	
Parry and Small (2009)		✓	✓			✓			✓	✓		
Calthrop et al. (2010)					✓							
Basso et al. (2011a)		✓	✓						✓	✓	✓	
Basso and Jara-Díaz (2012)		✓							✓		✓	

(*continued on next page*)

Table A.3 (*continued*)

	Operator cost function								Tech. details			
	Demand dependent	Frequency dependent	Vehicle size dependent	Fleet size dependent	Just infrastructure cost	Temporal cost interactions	Spatial cost interactions	Costly public funds (MCPF)	Endogenous cycle time	Interaction with congestion	Vehicle capacity constraint	Bus stop congestion
Basso and Silva (2014)		✓	✓	✓		✓		✓	✓	✓	✓	✓
Tirachini et al. (2014b)	✓	✓	✓				✓		✓	✓		✓
Tirachini (2014)	✓	✓							✓			✓
De Borger and Proost (2015)	✓	✓	✓						✓			
de Palma et al. (2015)*												
Fielbaum et al. (2016)			✓	✓					✓		✓	
Gschwender et al. (2016)			✓	✓					✓		✓	
Börjesson et al. (2017)		✓	✓	✓		✓				✓		
de Palma et al. (2017)		✓	✓					✓				
Hörcher and Graham (2018)		✓	✓				✓					
Hörcher et al. (2020b)			✓					✓				

*: *Papers that do not model operator costs and technological features explicitly.*

Table A.4

Decision variables in supply optimisation

	Fare level	Fare structure	Frequency	Customised scheduling	Fleet size	Vehicle size	Stop density	Line density	Line structure	Seat provision	Bus lane/ priorities	Fare collection technology	Number & operation of doors
Mohring (1972)			✓		✓		✓						
Glaister (1974)	✓												
Jansson (1980)			✓		✓	✓							
Kocur and Hendrickson (1982)	✓		✓					✓					
Carbajo (1988)	✓	✓											
Oldfield and Bly (1988)	✓				✓	✓							
Chang and Schonfeld (1991)	✓		✓					✓					
Kraus (1991)	✓												
Jansson (1993)	✓		✓										
De Borger et al. (1996)	✓												
De Borger and Wouters (1998)	✓		✓		✓								
Gronau (2000)			✓			✓							
Huang (2000)	✓												
Parry and Bento (2001)	✓												
Kraus and Yoshida (2002)	✓			✓	✓	✓							
Rietveld et al. (2002)	✓		✓			✓							
Jara-Díaz and Gschwender (2003a,b)			✓			✓							
Small (2004)			✓					✓					
Pels and Verhoef (2007)	✓		✓			✓							
Rietveld and van Woudenberg (2007)	✓		✓			✓							
Proost and Van Dender (2008)	✓		✓										
Ahn (2009)	✓		✓										
Jara-Díaz and Gschwender (2009a,b)	✓		✓		✓	✓							
Parry and Small (2009)	✓												
Calthrop et al. (2010)*													
Basso et al. (2011a)	✓		✓			✓	✓				✓		
Basso and Jara-Díaz (2012)	✓		✓										
Basso and Silva (2014)	✓		✓			✓	✓				✓		
Tirachini et al. (2014b)	✓		✓			✓				✓		✓	✓
Tirachini (2014)			✓				✓				✓	✓	
De Borger and Proost (2015)	✓		✓			✓							
de Palma et al. (2015)	✓	✓		✓						✓			
Fielbaum et al. (2016)			✓		✓				✓				
Gschwender et al. (2016)			✓		✓				✓				
Börjesson et al. (2017)	✓		✓			✓					✓		
de Palma et al. (2017)	✓	✓	✓	✓						✓			
Hörcher and Graham (2018)	✓		✓			✓							
Hörcher et al. (2020b)	✓		✓										

**: A policy appraisal model that does not perform a formal optimisation.*

References

Aapaoja, A., Eckhardt, J., Nykänen, L., 2017. Business models for MaaS. In: 1st International Conference on Mobility as a Service.

Abe, R., 2019. Introducing autonomous buses and taxis: quantifying the potential benefits in Japanese transportation systems. Transport. Res. Part A Pol. Pract. 126, 94–113.

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

Ainsalu, J., Arffman, V., Bellone, M., Ellner, M., Haapamäki, T., Haavisto, N., Josefson, E., Ismailogullari, A., Lee, B., Madland, O., et al., 2018. State of the art of automated buses. Sustainability 10 (9), 3118.

Ait Bihi Ouali, L., Graham, D.J., Barron, A., Trompet, M., 2020. Gender differences in the perception of safety in public transport. J. Roy. Stat. Soc. 183 (3), 737–769.

Alfa, A.S., Chen, M., 1995. Temporal distribution of public transport demand during the peak period. Eur. J. Oper. Res. 83, 137–153, 1984.

Allen, J., Muñoz, J.C., de Dios Ortúzar, J., 2019. On the Effect of Operational Service Attributes on Transit Satisfaction. Transportation, pp. 1–30.

Amirgholy, M., Gonzales, E.J., 2016. Demand responsive transit systems with time-dependent demand: User equilibrium, system optimum, and management strategy. Transp. Res. Part B Methodol. 92, 234–252.

Anas, A., 2013. A summary of the applications to date of RELU-TRAN, a microeconomic urban computable general equilibrium model. Environ. Plann. Plann. Des. 40 (6), 959–970.

Anas, A., Kim, I., 1996. General equilibrium models of polycentric urban land use with endogenous congestion and job agglomeration. J. Urban Econ. 40 (2), 232–256.

Anas, A., Liu, Y., 2007. A regional economy, land use, and transportation model (RELU-TRAN©): formulation, algorithm design, and testing. J. Reg. Sci. 47 (3), 415–455.

Anderson, S.P., De Palma, A., Thisse, J.-F., 1992. Discrete Choice Theory of Product Differentiation. MIT press.

Anupriya, Graham, D.J., Carbo, J.M., Anderson, R.J., Bansal, P., 2020a. Understanding the costs of urban rail transport operations. Transp. Res. Part B Methodol. 138, 292–316.

Anupriya, Graham, D.J., Hörcher, D., Anderson, R.J., Bansal, P., 2020b. Quantifying the ex-post causal impact of differential pricing on commuter trip scheduling in Hong Kong. Transport. Res. Pol. Pract. 141, 16–34.

Arnott, R., 1996. Taxi travel should be subsidized. J. Urban Econ. 40 (3), 316–333.

Arnott, R., 2007. Congestion tolling with agglomeration externalities. J. Urban Econ. 2 (62), 187–203.

Arnott, R., de Palma, A., Lindsey, R., 1990. Economics of a bottleneck. J. Urban Econ. 27 (1), 111–130.

Auriol, E., Warlters, M., 2012. The marginal cost of public funds and tax reform in Africa. J. Dev. Econ. 97 (1), 58–72.

Badia, H., Estrada, M., Robusté, F., 2014. Competitive transit network design in cities with radial street patterns. Transp. Res. Part B Methodol. 59, 161–181.

Bansal, P., Hurtubia, R., Tirachini, A., Daziano, R.A., 2019. Flexible estimates of heterogeneity in crowding valuation in the New York City subway. J. Choice Model. 31, 124–140.

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

Basso, L.J., Guevara, C.A., Gschwender, A., Fuster, M., 2011a. Congestion pricing, transit subsidies and dedicated bus lanes: efficient and practical solutions to congestion. Transport Pol. 18 (5), 676–684.

Basso, L.J., Jara-Díaz, S.R., 2006. Distinguishing multiproduct economies of scale from economies of density on a fixed-size transport network. Network. Spatial Econ. 6 (2), 149–162.

Basso, L.J., Jara-Díaz, S.R., 2012. Integrating congestion pricing, transit subsidies and mode choice. Transport. Res. Part A Pol. Pract. 46 (6), 890–900.

Basso, L.J., Jara-Díaz, S.R., Waters, W.G., 2011b. 12 Cost Functions for Transport Firms. A Handbook of Transport Economics, p. 273.

Basso, L.J., Silva, H.E., 2019. Optimal departure from marginal cost pricing in public transport. In: ITEA Annual Conference 2019, Paris.

Basu, D., Hunt, J.D., 2012. Valuing of attributes influencing the attractiveness of suburban train service in Mumbai city: A stated preference approach. Transport. Res. Part A Pol. Pract. 46 (9), 1465–1476.

Batarce, M., 2016. Estimation of urban bus transit marginal cost without cost data. Transp. Res. Part B Methodol. 90, 241–262.

Beaudry, A., Laporte, G., Melo, T., Nickel, S., 2010. Dynamic transportation of patients in hospitals. Spectrum 32 (1), 77–107.

Becker, G.S., 1965. A theory of the allocation of time. Econ. J. 493–517.

Beesley, M., Glaister, S., 1985a. Deregulating the bus industry in Britain: a reply. Transport Rev. 5.

Beesley, M., Glaister, S., 1985b. Deregulating the bus industry in Britain—(c) a response. Transport Rev. 5 (2), 133–142.

Bekhor, S., Dobler, C., Axhausen, K.W., 2011. Integration of activity-based and agent-based models: Case of Tel Aviv, Israel. Transport. Res. Rec. 2255 (1), 38–47.

Berechman, J., 1993. Public Transit Economics and Deregulation Policy. Elsevier.

Bergantino, A.S., Capozza, C., Capurso, M., 2015. The impact of open access on intra-and inter-modal rail competition. A national level analysis in Italy. Transport Pol. 39, 77–86.

Berggren, U., Johnsson, C., Svensson, H., Wretstand, A., 2019. Exploring waiting times in public transport through a semi-automated dedicated smartphone app survey. Travel Behav. Soc. 15, 1–14.

Beria, P., Redondi, R., Malighetti, P., 2016. The effect of open access competition on average rail prices. The case of Milan–Ancona. J. Rail Transport Plann. Manag. 6 (3), 271–283.

Berrebi, S.J., Watkins, K.E., Laval, J.A., 2015. A real-time bus dispatching policy to minimize passenger wait on a high frequency route. Transp. Res. Part B Methodol. 81, 377–389.

Bly, P., Oldfield, R., 1986. An analytic assessment of subsidies to bus services. Transport. Sci. 20 (3), 200–212.

Borck, R., Wrede, M., 2005. Political economy of commuting subsidies. J. Urban Econ. 57 (3), 478–499.

Borck, R., Wrede, M., 2008. Commuting subsidies with two transport modes. J. Urban Econ. 63 (3), 841–848.

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

Börjesson, M., 2012. Valuing perceived insecurity associated with use of and access to public transport. Transport Pol. 22, 1–10.

Börjesson, M., Eliasson, J., Rubensson, I., 2020. Distributional effects of public transport subsidies. J. Transport Geogr. 84, 102674.

Börjesson, M., Fung, C.M., Proost, S., 2017. Optimal prices and frequencies for buses in Stockholm. Econ. Transport. 9, 20–36.

Börjesson, M., Fung, C.M., Proost, S., Yan, Z., 2019. Do small cities need more public transport subsidies than big cities? J. Transport Econ. Pol. 53 (4), 275–298.

Börjesson, M., Rubensson, I., 2019. Satisfaction with crowding and other attributes in public transport. Transport Pol. 79, 213–222.

Bös, D., 1978. Distributional effects of maximisation of passenger miles. J. Transport Econ. Pol. 322–329.

Bösch, P.M., Becker, F., Becker, H., Axhausen, K.W., 2018. Cost-based analysis of autonomous mobility services. Transport Pol. 64, 76–91.

Braeutigam, R.R., 1999. Learning about transport costs. *Essays in transportation Economics and policy: a Handbook in Honor of John R. Meyer*. Reg. Sci. Urban Econ. 3, 57–97.

Brake, J., Mulley, C., Nelson, J.D., Wright, S., 2007. Key lessons learned from recent experience with flexible transport services. Transport Pol. 14 (6), 458–466.

Bröcker, J., Mercenier, J., 2011. General equilibrium models for transportation economics. In: De Palma, A., Lindsey, R., Quinet, E., Vickerman, R. (Eds.), A Handbook of Transport Economics. Edward Elgar Publishing, pp. 21–45.

Broman, E., Eliasson, J., 2019. Welfare effects of open access competition on railway markets. Transport. Res. Part A Pol. Pract. 129, 72–91.

Brons, M., Givoni, M., Rietveld, P., 2009. Access to railway stations and its potential in increasing rail use. Transport. Res. Part A Pol. Pract. 43 (2), 136–149.

Brown, S.J., Sibley, D.S., 1986. The Theory of Public Utility Pricing. Cambridge University Press.

Brueckner, J.K., 2005. Transport subsidies, system choice, and urban sprawl. Reg. Sci. Urban Econ. 35 (6), 715–733.

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

Caiati, V., Rasouli, S., Timmermans, H., 2020. Bundling, pricing schemes and extra features preferences for mobility as a service: Sequential portfolio choice experiment. Transport. Res. Part A Pol. Pract. 131, 123–148.

Calthrop, E., De Borger, B., Proost, S., 2010. Cost-benefit analysis of transport investments in distorted economies. Transport. Res. Part B 44 (7), 850–869.

Cantos, P., Pastor, J.M., Serrano, L., 1999. Productivity, efficiency and technical change in the European railways: a non-parametric approach. Transportation 26 (4), 337–357.

Cantos, P., Pastor, J.M., Serrano, L., 2010. Vertical and horizontal separation in the European railway sector and its effects on productivity. J. Transport Econ. Pol. 44 (2), 139–160.

Carbajo, J.C., 1988. The economics of travel passe: Non-uniform pricing in transport. J. Transport Econ. Pol. 22 (2), 153–173.

Cats, O., 2013. Multi-agent transit operations and assignment model. Procedia Comput. Sci. 19, 809–814.

Caves, D.W., Christensen, L.R., Tretheway, M.W., 1984. Economies of density versus economies of scale: why trunk and local service airline costs differ. Rand J. Econ. 15 (4), 471–489.

Cepeda, M., Cominetti, R., Florian, M., 2006. A frequency-based assignment model for congested transit networks with strict capacity constraints: characterization and computation of equilibria. Transport. Res. Part B 40 (6), 437–459.

Cervero, R., 1998. The Transit Metropolis: A Global Inquiry. Island Press.

Cervero, R., 2004. Transit-oriented Development in the United States: Experiences, Challenges, and Prospects. Transportation Research Board.

Chalabianlou, R., Lawrence, A., Baxter, B., 2015. A review and assessment of fare capping as a passenger incentive mechanism for Australia and New Zealand. In: Australasian Transport Research Forum (ATRF), 37th.

Chang, S., Schonfeld, P.M., 1991. Multiple period optimization of bus transit systems. Transp. Res. Part B Methodol. 25 (6), 453–478.

Chien, S.I., Qin, Z., 2004. Optimization of bus stop locations for improving transit accessibility. Transport. Plann. Technol. 27 (3), 211–227.

Cordeau, J.-F., Laporte, G., 2007. The dial-a-ride problem: models and algorithms. Ann. Oper. Res. 153 (1), 29–46.

Cortés, C., Jara-Díaz, S., Tirachini, A., 2011. Integrating short turning and deadheading in the optimization of transit services. Transport. Res. Part A 45 (5), 419–434.

Cowie, J., 2002. The production economics of a vertically separated railway: the case of the British train operating companies. Trasporti Europei 20.

Currie, G., 2005. The demand performance of bus rapid transit. J. Publ. Transport. 8 (1), 3.

Daganzo, C.F., 2010. Structure of competitive transit networks. Transp. Res. Part B Methodol. 44 (4), 434–446.

Daganzo, C.F., 2012. On the design of public infrastructure systems with elastic demand. Transp. Res. Part B Methodol. 46 (9), 1288–1293.

Danielis, R., Marcucci, E., 2002. Bottleneck road congestion pricing with a competing railroad service. Transport. Res. Part E 38 (5), 379–388.

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., Fosgerau, M., 2012. Information provision by regulated public transport companies. Transp. Res. Part B Methodol. 46 (4), 492–510.

De Borger, B., Kerstens, K., Costa, A., 2002. Public transit performance: What does one learn from frontier studies? Transport Rev. 22 (1), 1–38.

De Borger, B., Mayeres, I., Proost, S., Wouters, S., 1996. Optimal pricing of urban passenger transport: a simulation exercise for Belgium. J. Transport Econ. Pol. 30 (1), 31–54.

De Borger, B., Proost, S., 2015. The political economy of public transport pricing and supply decisions. Econ. Transport. 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 Publ. Finance 23 (5), 934–959.

De Borger, B., Wouters, S., 1998. Transport externalities and optimal pricing and supply decisions in urban transportation: a simulation analysis for Belgium. Reg. Sci. Urban Econ. 28 (2), 163–197.

de Jonge, B., Teunter, R.H., 2013. Optimizing itineraries in public transportation with walks between rides. Transp. Res. Part B Methodol. 55, 212–226.

de Palma, A., Kilani, M., Proost, S., 2015. Discomfort in mass transit and its implication for scheduling and pricing. Transp. Res. Part B Methodol. 71, 1–18.

de Palma, A., Lindsey, R., 2001. Optimal timetables for public transportation. Transp. Res. Part B Methodol. 35 (8), 789–813.

de Palma, A., Lindsey, R., 2007. Transport user charges and cost recovery. Res. Transport. Econ. 19, 29–57.

de Palma, A., Lindsey, R., Monchambert, G., 2017. The economics of crowding in rail transit. J. Urban Econ. 101, 106–122.

Delbosc, A., Currie, G., 2012. Modelling the causes and impacts of personal safety perceptions on public transport ridership. Transport Pol. 24, 302–309.

Delle Site, P., Filippi, F., 1998. Service optimization for bus corridors with short-turn strategies and variable vehicle size. Transport. Res. Part A 32 (1), 19–38.

Desaulniers, G., Hickman, M., 2007. Public transit. In: Barnhart, C., Laporte, G. (Eds.), Handbooks in Operations Research and Management Science, vol. 14. Elsevier, pp. 69–127. Transportation, chapter 2.

DeSerpa, A.C., 1971. A theory of the economics of time. Econ. J. 81 (324), 828–846.

Di Pietrantonio, L., Pelkmans, J., 2004. The economics of EU railway reform. J. Netw. Ind. 5 (3–4), 295–346.

Dodgson, J.S., Topham, N., 1987. Benefit-cost rules for urban transit subsidies: an integration of allocational, distributional and public finance issues. J. Transport Econ. Pol. 21 (1), 57–71.

El-Geneidy, A.M., Tetreault, P., Surprenant-Legault, J., 2010. Pedestrian access to transit: Identifying redundancies and gaps using a variable service area analysis. Transportation Research Board, 89th Annual Meeting, Washington DC.

Else, P.K., 1985. Optimal pricing and subsidies for scheduled transport services. J. Transport Econ. Pol. 19 (3), 263–279.

Errico, F., Crainic, T.G., Malucelli, F., Nonato, M., 2013. A survey on planning semi-flexible transit systems: methodological issues and a unifying framework. Transport. Res. C Emerg. Technol. 36, 324–338.

Evans, A., 1987. A theoretical comparison of competition with other economic regimes for bus services. J. Transport Econ. Pol. 7–36.

Evans, A., 1990. Competition and the structure of local bus markets. J. Transport Econ. Pol. 255–281.

Fan, W., Machemehl, R.B., 2009. Do transit users just wait for buses or wait with strategies? Some numerical results that transit planners should see. Transport. Res. Rec. 2111, 169–176.

Farahani, R.Z., Miandoabchi, E., Szeto, W.Y., Rashidi, H., 2013. A review of urban transportation network design problems. Eur. J. Oper. Res. 229 (2), 281–302.

Farber, S., Bartholomew, K., Li, X., Páez, A., Habib, K.M.N., 2014. Assessing social equity in distance based transit fares using a model of travel behavior. Transport. Res. Part A Pol. Pract. 67, 291–303.

Fernández, R., Planzer, R., 2002. On the capacity of bus transit systems. Transport Rev. 22 (3), 267–293.

Fielbaum, A., 2019. Strategic public transport design using autonomous vehicles and other new technologies. Int. J. Intell. Transport. Syst. Res. 1–9.

Fielbaum, A., Jara-Díaz, S., Gschwender, A., 2016. Optimal public transport networks for a general urban structure. Transp. Res. Part B Methodol. 94, 298–313.

Fielbaum, A., Jara-Díaz, S., Gschwender, A., 2017. A parametric description of cities for the normative analysis of transport systems. Network. Spatial Econ. 17 (2), 343–365.

Fielbaum, A., Jara-Díaz, S., Gschwender, A., 2018. Transit line structures in a general parametric city: the role of heuristics. Transport. Sci. 52 (5), 1092–1105.

Fielbaum, A., Jara-Díaz, S., Gschwender, A., 2020. Beyond the mohring effect: scale economies induced by transit lines structures design. Econ. Transport. 22, 100163.

FitzRoy, F., Smith, I., 1998. Public transport demand in Freiburg: why did patronage double in a decade? Transport Pol. 5 (3), 163–173.

FitzRoy, F., Smith, I., 1999. Season tickets and the demand for public transport. Kyklos 52 (2), 219–238.

Fosgerau, M., 2009. The marginal social cost of headway for a scheduled service. Transp. Res. Part B Methodol. 43 (8–9), 813–820.

Fosgerau, M., Engelson, L., 2011. The value of travel time variance. Transp. Res. Part B Methodol. 45 (1), 1–8.

Frankena, M.W., 1983. The efficiency of public transport objectives and subsidy formulas. J. Transport Econ. Pol. 17 (1), 67–76.

Fröidh, O., Byström, C., 2013. Competition on the tracks–Passengers' response to deregulation of interregional rail services. Transport. Res. Part A Pol. Pract. 56, 1–10.

Furth, P., Rahbee, A., 2000. Optimal bus stop spacing through dynamic programming and geographic modeling. Transport. Res. Rec. 1731, 15–22.

Furth, P.G., Mekuria, M.C., SanClemente, J.L., 2007. Stop spacing analysis using geographic information system tools with parcel and street network data. Transport. Res. Rec. 73–81, 2034.

Gagnepain, P., Ivaldi, M., 2002. Incentive regulatory policies: the case of public transit systems in France. Rand J. Econ. 605–629.

Gagnepain, P., Ivaldi, M., 2017. Economic efficiency and political capture in public service contracts. J. Ind. Econ. 65 (1), 1–38.

Gagnepain, P., Ivaldi, M., Martimort, D., 2013. The cost of contract renegotiation: evidence from the local public sector. Am. Econ. Rev. 103 (6), 2352–2383.

Gagnepain, P., Ivaldi, M., Vibes, C., 2011. The industrial organization of competition in local bus services. In: De Palma, A., Lindsey, R., Quinet, E., Vickerman, R. (Eds.), A Handbook of Transport Economics. Edward Elgar Publishing.

Garaix, T., Artigues, C., Feillet, D., Josselin, D., 2011. Optimization of occupancy rate in dial-a-ride problems via linear fractional column generation. Comput. Oper. Res. 38 (10), 1435–1442.

García-Ferrer, A., Bujosa, M., de Juan, A., Poncela, P., 2006. Demand forecast and elasticities estimation of public transport. J. Transport Econ. Pol. 40 (1), 45–67.

Garcia-Martinez, A., Cascajo, R., Jara-Díaz, S.R., Chowdhury, S., Monzon, A., 2018. Transfer penalties in multimodal public transport networks. Transport. Res. Part A Pol. Pract. 114, 52–66.

Gathon, H.-J., Pestieau, P., 1995. Decomposing efficiency into its managerial and its regulatory components: The case of European railways. Eur. J. Oper. Res. 80 (3), 500–507.

Gentile, G., Florian, M., Hamdouch, Y., Cats, O., Nuzzolo, A., 2016a. The theory of transit assignment: basic modelling frameworks. In: Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems. Springer, pp. 287–386.

Gentile, G., Noekel, K., Schmöcker, J.-D., Trozzi, V., Chandakas, E., 2016b. The theory of transit assignment: demand and supply phenomena. In: Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems. Springer, pp. 387–481.

Gibson, J., Baeza, I., Willlumsen, L.G., 1989. Bus-stops, congestion and congested bus-stops. Traffic Eng. Contr. 30 (6), 291–302.

Gkiotsalitis, K., Van Berkum, E., 2020. An exact method for the bus dispatching problem in rolling horizons. Transport. Res. C Emerg. Technol. 110, 143–165.

Gkritza, K., Karlaftis, M.G., Mannering, F.L., 2011. Estimating multimodal transit ridership with a varying fare structure. Transport. Res. Part A Pol. Pract. 45 (2), 148–160.

Glaister, S., 1974. Generalised consumer surplus and public transport pricing. Econ. J. 84 (336), 849–867.

Glaister, S., 1986. Bus deregulation, competition and vehicle size. J. Transport Econ. Pol. 217–244.

Glaister, S., Collings, J.J., 1978. Maximisation of passenger miles in theory and practice. J. Transport Econ. Pol. 304–321.

Gómez-Lobo, A., Briones, J., 2014. Incentives in bus concession contracts: a review of several experiences in Latin america. Transport Rev. 34 (2), 246–265.

Gronau, R., 2000. Optimum diversity in the public transport market. J. Transport Econ. Pol. 21–41.

Growitsch, C., Wetzel, H., 2009. Testing for economies of scope in European railways. J. Transport Econ. Pol. 1–24.

Gschwender, A., Jara-Díaz, S., Bravo, C., 2016. Feeder-trunk or direct lines? Economies of density, transfer costs and transit structure in an urban context. Transport. Res. Part A Pol. Pract. 88, 209–222.

Guidon, S., Wicki, M., Bernauer, T., Axhausen, K., 2020. Transportation service bundling–For whose benefit? Consumer valuation of pure bundling in the passenger transportation market. Transport. Res. Part A Pol. Pract. 131, 91–106.

Guihaire, V., Hao, J.-K., 2008. Transit network design and scheduling: A global review. Transport. Res. Part A Pol. Pract. 42 (10), 1251–1273.

Gwilliam, K., 2008. A review of issues in transit economics. Res. Transport. Econ. 23 (1), 4–22.

Gwilliam, K., Nash, C., Mackie, P., 1985a. Deregulating the bus industry in Britain: a rejoinder. Transport Rev. 5 (3), 215–222.

Gwilliam, K.M., Nash, C., Mackie, P., 1985b. Deregulating the bus industry in Britain—(b) the case against. Transport Rev. 5 (2), 105–132.

Häll, C.H., Andersson, H., Lundgren, J.T., Värbrand, P., 2009. The integrated dial-a-ride problem. Publ. Transport 1 (1), 39–54.

Hall, J.D., Palsson, C., Price, J., 2018. Is Uber a substitute or complement for public transit? J. Urban Econ. 108, 36–50.

Harris, N.G., 2006. Train boarding and alighting rates at high passenger loads. J. Adv. Transport. 40 (3), 249–263.

Harris, N.G., Anderson, R.J., 2007. An international comparison of urban rail boarding and alighting rates. Proc. Inst. Mech. Eng. - Part F J. Rail Rapid Transit 221 (4), 521–526.

Haywood, L., Koning, M., Monchambert, G., 2017. Crowding in public transport: Who cares and why? Transport. Res. Part A Pol. Pract. 100, 215–227.

Hensher, D., 2017a. Future bus transport contracts under a mobility as a service (MaaS) regime in the digital age: are they likely to change? Transport. Res. Part A 98, 86–96.

Hensher, D., Golob, T., 2008. Bus rapid transit systems: a comparative assessment. Transportation 35 (4), 501–518.

Hensher, D.A., 2007. Bus Transport: Economics, Policy and Planning. Research in Transportation Economics, vol. 18. Elsevier.

Hensher, D.A., 2017b. Public service contracts: the economics of reform with special reference to the bus sector. In: The Routledge Handbook of Transport Economics. Routledge, pp. 91–107.
Hensher, D.A., Stanley, J., 2010. Contracting regimes for bus services: what have we learnt after 20 years? Res. Transport. Econ. 29 (1), 140–144.
Hickman, M., Blume, K., 2001. Modeling cost and passenger level of service for integrated transit service. In: Computer-aided Scheduling of Public Transport. Springer, pp. 233–251.
Ho, C.Q., Mulley, C., Hensher, D.A., 2020. Public preferences for mobility as a service: Insights from stated preference surveys. Transport. Res. Part A Pol. Pract. 131, 70–90.
Ho, S.C., Szeto, W., Kuo, Y.-H., Leung, J.M., Petering, M., Tou, T.W., 2018. A survey of dial-a-ride problems: literature review and recent developments. Transp. Res. Part B Methodol. 111, 395–421.
Holmgren, J., 2007. Meta-analysis of public transport demand. Transport. Res. Part A Pol. Pract. 41 (10), 1021–1035.
Hörcher, D., 2018. The Economics of Crowding in Urban Rail Transport. PhD thesis. Imperial College London.
Hörcher, D., De Borger, B., Graham, D.J., 2020a. Benefit Spillovers and Subsidy Exporting in Inter-regional Public Transport Provision. Available at: SSRN 3710035.
Hörcher, D., De Borger, B., Seifu, W., Graham, D.J., 2020b. 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. Part B Methodol. 108, 106–126.
Hörcher, D., Graham, D.J., 2020a. The Gini Index of Demand Imbalances in Public Transport. Transportation, pp. 1–24.
Hörcher, D., Graham, D.J., 2020b. MaaS economics: should we fight car ownership with subscriptions to alternative modes? Econ. Transport. 22, 100167.
Hörcher, D., Graham, D.J., Anderson, R.J., 2017. Crowding cost estimation with large scale smart card and vehicle location data. Transp. Res. Part B Methodol. 95, 105–125.
Hörcher, D., Graham, D.J., Anderson, R.J., 2018a. The economic inefficiency of travel passes under crowding externalities and endogenous capacity. J. Transport Econ. Pol. 52 (1), 1–22.
Hörcher, D., Graham, D.J., Anderson, R.J., 2018b. The economics of seat provision in public transport. Transport. Res. E Logist. Transport. Rev. 109, 277–292.
Hörcher, D., Singh, R., Graham, D., 2020c. Social Distancing in Public Transport: Mobilising New Technologies for Demand Management under the Covid-19 Crisis. Technical report, Available at: SSRN.
Horni, A., Nagel, K., W Axhausen, K., 2016. The Multi-Agent Transport Simulation MATSim. Ubiquity Press.
Huang, H.-J., 2000. Fares and tolls in a competitive system with transit and highway: the case with two groups of commuters. Transport. Res. Part E 36 (4), 267–284.
Huang, H.-J., 2002. Pricing and logit-based mode choice models of a transit and highway system with elastic demand. Eur. J. Oper. Res. 140 (3), 562–570.
Huang, H.-J., Tian, Q., Gao, Z.-Y., 2005. An equilibrium model in urban transit riding and fare polices. Algorithmic Appl. Manag. 3521, 112–121.
Ibarra-Rojas, O., Delgado, F., Giesen, R., M, J.C., 2015. Planning, operation, and control of bus transport systems: A literature review. Transp. Res. Part B Methodol. 77, 38–75.
Ingvardson, J.B., Nielsen, O.A., Raveau, S., Nielsen, B.F., 2018. Passenger arrival and waiting time distributions dependent on train service frequency and station characteristics: a smart card data analysis. Transport. Res. C Emerg. Technol. 90, 292–306.
Iossa, E., Martimort, D., 2011. The theory of incentives applied to the transport sector. In: De Palma, A., Lindsey, R., Quinet, E., Vickerman, R. (Eds.), A Handbook of Transport Economics. Edward Elgar Publishing.
Iseki, H., Taylor, B.D., 2009. Not all transfers are created equal: Towards a framework relating transfer connectivity to travel behaviour. Transport Rev. 29 (6), 777–800.
Iseki, H., Taylor, B.D., 2010. The demographics of public transit subsidies: A case study of Los Angeles. Technical report. University of California Transportation Center. Retrieved from. https://escholarship.org/uc/item/9nq526f1.
Jackson, R., 1975. Optimal subsidies for public transit. J. Transport Econ. Pol. 3–15.
Jansson, J., 1979. Marginal cost pricing of scheduled transport services. J. Transport Econ. Pol. 13 (3), 268–294.
Jansson, J.O., 1980. A simple bus line model for optimisation of service frequency and bus size. J. Transport Econ. Pol. 53–80.
Jansson, J.O., Holmgren, J., Ljungberg, A., 2015. Pricing public transport services. In: Handbook of Research Methods and Applications in Transport Economics and Policy. Springer, pp. 101–122.
Jansson, K., 1993. Optimal public transport price and service frequency. J. Transport Econ. Pol. 27 (1), 33–50.
Jara-Díaz, S., Gschwender, A., 2005. Making pricing work in public transport provision. In: Button, K.J., Hensher, D.A. (Eds.), Handbook of Transport Strategy, Policy and Institutions. Elsevier Ltd, Amsterdam, pp. 447–459.
Jara-Díaz, S., Gschwender, A., 2009a. The effect of financial constraints on the optimal design of public transport services. Transportation 36 (1), 65–75.
Jara-Díaz, S., 2007. Transport Economic Theory. Emerald Group Publishing Limited.
Jara-Díaz, S., Cruz, D., Casanova, C., 2016. Optimal pricing for travelcards under income and car ownership inequities. Transport. Res. Part A Pol. Pract. 94, 470–482.
Jara-Díaz, S., Fielbaum, A., Gschwender, A., 2017. Optimal fleet size, frequencies and vehicle capacities considering peak and off-peak periods in public transport. Transport. Res. Part A Pol. Pract. 106, 65–74.
Jara-Díaz, S., Fielbaum, A., Gschwender, A., 2020. Strategies for transit fleet design considering peak and off-peak periods using the single-line model. Transp. Res. Part B Methodol. 142, 1–18.
Jara-Díaz, S., Gschwender, A., 2003a. Towards a general microeconomic model for the operation of public transport. Transport Rev. 23 (4), 453–469.
Jara-Díaz, S., Gschwender, A., Bravo, C., 2018. Total cost minimizing transit route structures considering trips towards CBD and periphery. Transportation 45 (6), 1701–1720.
Jara Díaz, S.R., 1982. The estimation of transport cost functions: a methodological review. Transport Rev. 2 (3), 257–278.
Jara-Díaz, S.R., Basso, L.J., 2003. Transport cost functions, network expansion and economies of scope. Transport. Res. E Logist. Transport. Rev. 39 (4), 271–288.
Jara-Díaz, S.R., Cortés, C.E., 1996. On the calculation of scale economies from transport cost functions. J. Transport Econ. Pol. 157–170.
Jara-Díaz, S.R., Gschwender, A., 2003b. From the single line model to the spatial structure of transit services: corridors or direct? J. Transport Econ. Pol. 37 (2), 261–277.
Jara-Díaz, S.R., Gschwender, A., 2009b. The effect of financial constraints on the optimal design of public transport services. Transportation 36 (1), 65–75.
Jara-Díaz, S.R., Gschwender, A., Ortega, M., 2012. Is public transport based on transfers optimal? A theoretical investigation. Transp. Res. Part B Methodol. 46 (7), 808–816.
Jara-Díaz, S.R., Gschwender, A., Ortega, M., 2014. The impact of a financial constraint on the spatial structure of public transport services. Transportation 41 (1), 21–36.
Kaddoura, I., Kickhöfer, B., Neumann, A., Tirachini, A., 2014. Agent-based optimisation of public transport supply and pricing: impacts of activity scheduling decisions and simulation randomness. Transportation 42 (6), 1039–1061.
Kaddoura, I., Kickhöfer, B., Neumann, A., Tirachini, A., 2015. Optimal public transport pricing: towards an agent-based marginal social cost approach. J. Transport Econ. Pol. 49 (2), 200–218 forthcoming.
Kennedy, D., 1995. London bus tendering: an overview. Transport Rev. 15 (3), 253–264.
Kepaptsoglou, K., Karlaftis, M., 2009. Transit route network design problem. J. Transport. Eng. 135 (8), 491–505.
Kerin, P.D., 1992. Efficient bus fares. Transport Rev. 12 (1), 33–47.
Kim, M.E., Schonfeld, P., 2014. Integration of conventional and flexible bus services with timed transfers. Transp. Res. Part B Methodol. 68, 76–97.
Kleven, H.J., Kreiner, C.T., 2006. The marginal cost of public funds: hours of work versus labor force participation. J. Publ. Econ. 90 (10–11), 1955–1973.
Kocur, G., Hendrickson, C., 1982. Design of local bus service with demand equilibration. Transport. Sci. 16 (2), 149–170.
Koffman, D., 2004. Operational Experiences with Flexible Transit Services. TCRP Synthesis Report, vol. 53. Transportation Research Board.
Kraus, M., 1991. Discomfort externalities and marginal cost transit fares. J. Urban Econ. 29 (2), 249–259.
Kraus, M., Yoshida, Y., 2002. The commuter's time-of-use decision and optimal pricing and service in urban mass transit. J. Urban Econ. 51 (1), 170–195.
Kroes, E., Kouwenhoven, M., Debrincat, L., Pauget, N., 2014. Value of crowding on public transport in île-de-France, France. Transp. Res. Rec. J. Transport. Res. Board 2417, 37–45.
Krygsman, S., Dijst, M., Arentze, T., 2004. Multimodal public transport: an analysis of travel time elements and the interconnectivity ratio. Transport Pol. 11 (3), 265–275.
Laffont, J.-J., Tirole, J., 1993. A Theory of Incentives in Procurement and Regulation. MIT Press.
Leiva, C., Muñoz, J., Giesen, R., Larrain, H., 2010. Design of limited-stop services for an urban bus corridor with capacity constraints. Transport. Res. Part B 44 (10), 1186–1201.
Leland, H.E., Meyer, R.A., 1976. Monopoly pricing structures with imperfect discrimination. Bell J. Econ. 449–462.
LeRoy, S.F., Sonstelie, J., 1983. Paradise lost and regained: transportation innovation, income, and residential location. J. Urban Econ. 13 (1), 67–89.
Li, X., Quadrifoglio, L., 2010. Feeder transit services: choosing between fixed and demand responsive policy. Transport. Res. C Emerg. Technol. 18 (5), 770–780.
Li, Z., Hensher, D.A., 2011. Crowding and public transport: a review of willingness to pay evidence and its relevance in project appraisal. Transport Pol. 18 (6), 880–887.
Li, Z.-C., Huang, H.-J., Yang, H., 2020. Fifty years of the bottleneck model: a bibliometric review and future research directions. Transp. Res. Part B Methodol. 139, 311–342.
Lindsey, R., 2006. Do economists reach a conclusion on road pricing? The intellectual history of an idea. Econ. J. Watch 3 (2), 292–379.
Littlechild, S.C., 1975. Two-part tariffs and consumption externalities. Bell J. Econ. 661–670.
Liu, Y., Bunker, J., Ferreira, L., 2010. Transit users' route-choice modelling in transit assignment: a review. Transport Rev. 30 (6), 753–769.
Mackie, P., 1983. Competition in the bus industry. Publ. Money Manag. 3 (1), 45–48.
Mackie, P., Preston, J., Nash, C., 1995. Bus deregulation: ten years on. Transport Rev. 15 (3), 229–251.
Malucelli, F., Nonato, M., Pallottino, S., 1999. Demand adaptive systems: some proposals on flexible transit. In: Operational Research in Industry. Springer, pp. 157–182.
Matas, A., 2004. Demand and revenue implications of an integrated public transport policy: the case of madrid. Transport Rev. 24 (March 2013), 195–217.
Matas, A., Raymond, J.-L., Ruiz, A., 2020. Economic and distributional effects of different fare schemes: Evidence from the metropolitan region of Barcelona. Transport. Res. Part A Pol. Pract. 138, 1–14.
Matyas, M., Kamargianni, M., 2018. The potential of Mobility as a Service bundles as a mobility management tool. Transportation, pp. 1–18.
Mayeres, I., 2001. Equity and transport policy reform. *Energy, Transport and Environment Working Papers Series N14*. Katholieke Universiteit Leuven.
McCarthy, O.T., Caulfield, B., O'Mahony, M., 2016. How transport users perceive personal safety apps. Transport. Res. F Traffic Psychol. Behav. 43, 166–182.

McFadden, D., 1973. Conditional logit analysis of qualitative choice behaviour. In: Zarembka, P. (Ed.), Frontiers in Econometrics. Academic Press New York, New York, NY, USA, pp. 105–142.

Mizutani, F., Smith, A., Nash, C., Uranishi, S., 2015. Comparing the costs of vertical separation, integration, and intermediate organisational structures in European and East Asian railways. J. Transport Econ. Pol. 49 (3), 496–515.

Moccia, L., Giallombardo, G., Laporte, G., 2017. Models for technology choice in a transit corridor with elastic demand. Transp. Res. Part B Methodol. 104, 733–756.

Mogridge, M.J., 1997. The self-defeating nature of urban road capacity policy: a review of theories, disputes and available evidence. Transport Pol. 4 (1), 5–23.

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

Mohring, H., 1983. Minibuses in urban transportation. J. Urban Econ. 14 (3), 293–317.

Mohring, H., Harwitz, M., 1962. Highway Benefits: an Analytical Framework. Northwestern University Press, Evanston, IL.

Molenbruch, Y., Braekers, K., Caris, A., 2017. Typology and literature review for dial-a-ride problems. Ann. Oper. Res. 259 (1–2), 295–325.

Mulley, C., Nelson, J.D., 2009. Flexible transport services: a new market opportunity for public transport. Res. Transport. Econ. 25 (1), 39–45.

Mulley, C., Nelson, J.D., Wright, S., 2018. Community transport meets mobility as a service: On the road to a new a flexible future. Res. Transport. Econ. 69, 583–591.

Muñoz, J.C., Cortés, C.E., Giesen, R., Sáez, D., Delgado, F., Valencia, F., Cipriano, A., 2013. Comparison of dynamic control strategies for transit operations. Transport. Res. C Emerg. Technol. 28, 101–113.

Nash, C., 2011. Competition and regulation in rail transport. In: De Palma, A., Lindsey, R., Quinet, E., Vickerman, R. (Eds.), A Handbook of Transport Economics. Edward Elgar Publishing.

Nash, C.A., 1978. Management objectives, fares and service levels in bus transport. J. Transport Econ. Pol. 70–85.

Nash, C.A., 1982. Economics of Public Transport. Longman Group UK Limited.

Nash, C.A., 1993. British bus deregulation. Econ. J. 103 (419), 1042–1049.

Navidi, Z., Ronald, N., Winter, S., 2018. Comparison between ad-hoc demand responsive and conventional transit: a simulation study. Publ. Transport 10 (1), 147–167.

Nelson, J.D., Wright, S., Masson, B., Ambrosino, G., Naniopoulos, A., 2010. Recent developments in flexible transport services. Res. Transport. Econ. 29 (1), 243–248.

Newell, G.F., 1971. Dispatching policies for a transportation route. Transport. Sci. 5 (1), 91–105.

Newell, G.F., 1982. Applications of Queueing Theory, 2nd ed. Chapman and Hall, New York, NY.

Nourbakhsh, S.M., Ouyang, Y., 2012. A structured flexible transit system for low demand areas. Transp. Res. Part B Methodol. 46 (1), 204–216.

Oates, W., 1972. Fiscal Federalism. Harcourt Brace Jovanovich, New York.

Oi, W.Y., 1971. A Disneyland dilemma: two-part tariffs for a Mickey Mouse monopoly. Q. J. Econ. 77–96.

Oldfield, R., Bly, P., 1988. An analytic investigation of optimal bus size. Transp. Res. Part B Methodol. 22 (5), 319–337.

Ongel, A., Loewer, E., Roemer, F., Sethuraman, G., Chang, F., Lienkamp, M., 2019. Economic assessment of autonomous electric microtransit vehicles. Sustainability 11 (3).

Osuna, E., Newell, G., 1972. Control strategies for an idealized bus system. Transport. Sci. 6 (1), 52–71.

Oum, T.H., Waters II., W.G., 1996. A survey of recent developments in transportation cost function research. Logist. Transport Rev. 32 (4), 423.

Oum, T.H., Waters II., W.G., Yong, J.-S., 1992. Concepts of price elasticities of transport demand and recent empirical estimates: an interpretative survey. J. Transport Econ. Pol. 139–154.

Oum, T.H., Yu, C., 1994. Economic efficiency of railways and implications for public policy: A comparative study of the OECD countries' railways. J. Transport Econ. Pol. 121–138.

Parragh, S.N., Pinho de Sousa, J., Almada-Lobo, B., 2015. The dial-a-ride problem with split requests and profits. Transport. Sci. 49 (2), 311–334.

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.H., Small, K.A., 2009. Should urban transit subsidies be reduced? Am. Econ. Rev. 99 (3), 700–724.

Paulley, N., Balcombe, R., Mackett, R., Titheridge, H., Preston, J., Wardman, M., Shires, J., White, P., 2006. The demand for public transport: the effects of fares, quality of service, income and car ownership. Transport Pol. 13 (4), 295–306.

Peer, S., Knockaert, J., Verhoef, E.T., 2016. Train commuters' scheduling preferences: evidence from a large-scale peak avoidance experiment. Transp. Res. Part B Methodol. 83, 314–333.

Pels, E., Rietveld, P., 2007. Cost functions in transport. In: Handbook of Transport Modelling, second ed. Emerald Group Publishing Limited, pp. 381–394.

Pels, E., Verhoef, E., 2007. Infrastructure pricing and competition between modes in urban transport. Environ. Plann. 39, 2119–2138.

Ponti, M., 2011. Competition, regulation and public service obligations. In: De Palma, A., Lindsey, R., Quinet, E., Vickerman, R. (Eds.), A Handbook of Transport Economics. Edward Elgar Publishing.

Potts, J.F., Marshall, M.A., Crockett, E.C., Washington, J., 2010. A guide for planning and operating flexible public transportation services. TCRP Synthesis Report. Transportation Research Board, p. 140.

Preston, J., 1996. The economics of British Rail privatization: an assessment. Transport Rev. 16 (1), 1–21.

Preston, J., 2008. Competition in transit markets. Res. Transport. Econ. 23 (1), 75–84.

Proost, S., De Borger, B., Koskenoja, P., 2007. Public finance aspects of transport charging and investments. In: de Palma, A., Lindsey, R., Proost, S. (Eds.), Investment and the Use of Tax and Toll Revenues in the Transport Sector, Research in Transportation Economics, vol. 19. Elsevier, pp. 59–80 chapter 3.

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

Proost, S., Van Dender, K., 2008. Optimal urban transport pricing in the presence of congestion, economies of density and costly public funds. Transport. Res. Part A Pol. Pract. 42 (9), 1220–1230.

Quadrifoglio, L., Li, X., 2009. A methodology to derive the critical demand density for designing and operating feeder transit services. Transp. Res. Part B Methodol. 43 (10), 922–935.

Quinet, E., 2005. Alternative pricing doctrines. Res. Transport. Econ. 14, 19–47.

Raveau, S., Guo, Z., Muñoz, J.C., Wilson, N.H., 2014. A behavioural comparison of route choice on metro networks: time, transfers, crowding, topology and socio-demographics. Transport. Res. Part A Pol. Pract. 66, 185–195.

Reinhardt, L.B., Clausen, T., Pisinger, D., 2013. Synchronized dial-a-ride transportation of disabled passengers at airports. Eur. J. Oper. Res. 225 (1), 106–117.

Rietveld, P., Dekkers, J., van Woudenberg, S., 2002. Choice of frequency and vehicle size in rail transport. Implications for marginal external costs. Eur. J. Transport Infrastruct. Res. 2 (2), 95–111.

Rietveld, P., van Woudenberg, S., 2007. Second-best decision making of railway operators: how to fix fares, frequency and vehicle size. J. Transport Econ. Pol. 41, 363–385.

Robson, E.N., Wijayaratna, K.P., Dixit, V.V., 2018. A review of computable general equilibrium models for transport and their applications in appraisal. Transport. Res. Part A Pol. Pract. 116, 31–53.

Rubensson, I., Susilo, Y., Cats, O., 2020. Is flat fare fair? Equity impact of fare scheme change. Transport Pol. 91, 48–58.

Ruiz-Rúa, A., Palacín, R., 2013. Towards a liberalised European high speed railway sector: analysis and modelling of competition using game theory. Eur. Transport Res. Rev. 5 (1), 53–63.

Sasaki, K., 1990. Income class, modal choice, and urban spatial structure. J. Urban Econ. 27 (3), 322–343.

Savage, I., 1997. Scale economies in United States rail transit systems. Transport. Res. Part A Pol. Pract. 31 (6), 459–473.

Selviaridis, K., Wynstra, F., 2015. Performance-based contracting: A literature review and future research directions. Int. J. Prod. Res. 53 (12), 3505–3540.

Sherman, R., 1971. Congestion interdependence and urban transit fares. Econometrica 39 (3), 565–576.

Singh, R., Hörcher, D., Graham, D.J., Anderson, R.J., 2020. Decomposing journey times on urban metro systems via semiparametric mixed methods. Transport. Res. C Emerg. Technol. 114, 140–163.

Small, K., 2008. Urban transportation policy: a guide and road map. In: Inman, Robert P. (Ed.), The Urban Enigma: City Problems, City Prospects. Princeton University Press.

Small, K., Rosen, H., 1981. Applied welfare economics with discrete choice models. Econometrica 49 (1), 105–130.

Small, K.A., 2004. Road pricing and public transport. Res. Transport. Econ. 9 (1), 133–158.

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

Smith, G., Sochor, J., Karlsson, I.M., 2018. Mobility as a service: development scenarios and implications for public transport. Res. Transport. Econ. 69, 592–599.

Sochor, J., Arby, H., Karlsson, I.M., Sarasini, S., 2018. A topological approach to mobility as a service: a proposed tool for understanding requirements and effects, and for aiding the integration of societal goals. Res. Transport. Bus. Manag. 27, 3–14.

Sørensen, P.B., 2010. Swedish Tax Policy: Recent Trends and Future Challenges. Finansdepartementet, Regeringskansliet.

Su, Q., DeSalvo, J.S., 2008. The effect of transportation subsidies on urban sprawl. J. Reg. Sci. 48 (3), 567–594.

Sun, Y., Guo, Q., Schonfeld, P., Li, Z., 2016. Implications of the cost of public funds in public transit subsidization and regulation. Transport. Res. Part A Pol. Pract. 91, 236–250.

Tabuchi, T., 1993. Bottleneck congestion and modal split. J. Urban Econ. 34 (3), 414–431.

Tang, Y., Yang, H., Wang, B., Huang, J., Bai, Y., 2020. A Pareto-improving and revenue-neutral scheme to manage mass transit congestion with heterogeneous commuters. Transport. Res. C Emerg. Technol. 113, 245–259.

Tian, Q., Huang, H.-J., Yang, H., 2007. Equilibrium properties of the morning peak-period commuting in a many-to-one mass transit system. Transp. Res. Part B Methodol. 41 (6), 616–631.

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., 2014. The economics and engineering of bus stops: Spacing, design and congestion. Transport. Res. Part A Pol. Pract. 59, 37–57.

Tirachini, A., 2020. Ride-hailing, travel behaviour and sustainable mobility: an international review. Transportation 47, 2011–2047.

Tirachini, A., Antoniou, C., 2020. The economics of automated public transport: effects on operator cost, travel time, fare and subsidy. Econ. Transport. 21, 100151.

Tirachini, A., Cats, O., 2020. COVID-19 and public transportation: current assessment, prospects, and research needs. J. Publ. Transport. 22 (1), 1–21.

Tirachini, A., Hensher, D., Jara-Díaz, S., 2010a. Comparing operator and users costs of light rail, heavy rail and bus rapid transit over a radial public transport network. Res. Transport. Econ. 29 (1), 231–242.

Tirachini, A., Hensher, D., Jara-Díaz, S., 2010b. Restating modal investment priority with an improved model for public transport analysis. Transport. Res. Part E 46 (6), 1148–1168.

Tirachini, A., Hensher, D.A., 2011. Bus congestion, optimal infrastructure investment and the choice of a fare collection system in dedicated bus corridors. Transport. Res. Part B 45 (5), 828–844.

Tirachini, A., Hensher, D.A., 2012. Multimodal transport pricing: first best, second best and extensions to non-motorized transport. Transport Rev. 32 (2), 181–202.

Tirachini, A., Hensher, D.A., Bliemer, M.C., 2014a. Accounting for travel time variability in the optimal pricing of cars and buses. Transportation 41, 947–971.

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

Tirachini, A., Hensher, D.A., Rose, J.M., 2014b. Multimodal pricing and optimal design of urban public transport: the interplay between traffic congestion and bus crowding. Transp. Res. Part B Methodol. 61, 33–54.

Tirachini, A., Hurtubia, R., Dekker, T., Daziano, R.A., 2017. Estimation of crowding discomfort in public transport: results from Santiago de Chile. Transport. Res. Part A Pol. Pract. 103, 311–326.

Tirachini, A., Proost, S., 2021. Transport taxes and subsidies in developing countries: The effect of income inequality aversion. Econ. Transport. 25, 100206.

Tirachini, A., Sun, L., Erath, A., Chakirov, A., 2016. Valuation of sitting and standing in metro trains using revealed preferences. Transport Pol. 47, 94–104.

Tiznado-Aitken, I., Muñoz, J.C., Hurtubia, R., 2020. Who gains in a distance-based public transport fare scheme? Accessibility, urban form, and equity implications in santiago, Chile. In: Mulley, C., Nelson, J. (Eds.), Urban Form and Accessibility. Elsevier, pp. 265–288.

Toledo, T., Cats, O., Burghout, W., Koutsopoulos, H.N., 2010. Mesoscopic simulation for transit operations. Transport. Res. C Emerg. Technol. 18 (6), 896–908.

Transport Strategy Centre, 2020. How Are the World's Public Transport Providers and Airport Operators Responding to the Coronavirus (COVID-19)? Technical Report. Imperial College London Projects, 15 September 2020.

Tscharaktschiew, S., Hirte, G., 2012. Should subsidies to urban passenger transport be increased? A spatial CGE analysis for a German metropolitan area. Transport. Res. Part A Pol. Pract. 46 (2), 285–309.

Turvey, R., Mohring, H., 1975. Optimal bus fares. J. Transport Econ. Pol. 9 (3), 280–286.

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

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

Varela, J.M.L., Börjesson, M., Daly, A., 2018. Public transport: One mode or several? Transport. Res. Part A Pol. Pract. 113, 137–156.

Verhoef, E., Nijkamp, P., 2003. Externalities in the Urban Economy. Tinbergen Institute Discussion Paper Series, 078/3.

Vickrey, W.S., 1955. A proposal for revising New York's subway fare structure. J. Oper. Res. Soc. Am. 3 (1), 38–68.

Vickrey, W.S., 1963. Pricing in urban and suburban transport. Am. Econ. Rev. 53 (2), 452–465.

Vickrey, W.S., 1969. Congestion theory and transport investment. Am. Econ. Rev. 59 (2), 251–260.

Vigren, A., 2017. Competition in Swedish passenger railway: entry in an open access market and its effect on prices. Econ. Transport. 11, 49–59.

Wadud, Z., 2017. Fully automated vehicles: A cost of ownership analysis to inform early adoption. Transport. Res. Part A Pol. Pract. 101, 163–176.

Wahba, M., Shalaby, A., 2005. Multiagent learning-based approach to transit assignment problem: a prototype. Transport. Res. Rec. (1), 96–105, 1926.

Walters, A.A., 1982. Externalities in urban buses. J. Urban Econ. 11 (1), 60–72.

Wang, J.Y.T., Lindsey, R., Yang, H., 2011. Nonlinear pricing on private roads with congestion and toll collection costs. Transp. Res. Part B Methodol. 45 (1), 9–40.

Wardman, M., 2004. Public transport values of time. Transport Pol. 11 (4), 363–377.

Wardman, M., 2012. Review and meta-analysis of UK time elasticities of travel demand. Transportation 39 (3), 465–490.

Wardman, M., 2014. Price elasticities of surface travel demand a meta-analysis of UK evidence. J. Transport Econ. Pol. 48 (3), 367–384.

Wardman, M., Murphy, P., 2015. Passengers' valuations of train seating layout, position and occupancy. Transport. Res. Part A Pol. Pract. 74, 222–238.

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

Waters II., W.G., 2007. Evolution of railroad economics. Res. Transport. Econ. 20, 11–67.

Whelan, G., Johnson, D., 2004. Modelling the impact of alternative fare structures on train overcrowding. Int. J. Transport Manag. 2 (1), 51–58.

White, P.R., 1990. Bus deregulation: a welfare balance sheet. J. Transport Econ. Pol. 311–332.

Wong, Y.Z., Hensher, D.A., Mulley, C., 2018. Emerging Transport Technologies and the Modal Efficiency Framework: A Case for Mobility as a Service (MaaS). ITLS Working Papers.

Wrede, M., 2001. Should commuting expenses be tax deductible? A welfare analysis. J. Urban Econ. 49 (1), 80–99.

Yang, H., Tang, Y., 2018. Managing rail transit peak-hour congestion with a fare-reward scheme. Transp. Res. Part B Methodol. 110, 122–136.

Yap, M., Cats, O., van Arem, B., 2020. Crowding valuation in urban tram and bus transportation based on smart card data. Transportmetrica: Transport. Sci. 16 (1), 23–42.

Yoshida, Y., 2008. Commuter arrivals and optimal service in mass transit: does queuing behavior at transit stops matter? Reg. Sci. Urban Econ. 38 (3), 228–251.

Zhang, F., Yang, H., Liu, W., 2014. The Downs–Thomson Paradox with responsive transit service. Transport. Res. Part A Pol. Pract. 70, 244–263.

Zhang, W., Jenelius, E., Badia, H., 2019. Efficiency of semi-autonomous and fully autonomous bus services in trunk-and-branches networks. J. Adv. Transport. 2019, 7648735, 2019.