The industrial organization of competition in local bus services

Philippe Gagnepain, Marc Ivaldi, Catherine Vibes

To cite this version:

Philippe Gagnepain, Marc Ivaldi, Catherine Vibes. The industrial organization of competition in local bus services. De Palma, Lindsey, Quinet, et Vickerman. A handbook of transport economics, Edward Elgar, pp.744-762, 2011. <hal-00684161>

HAL Id: hal-00684161
https://hal.archives-ouvertes.fr/hal-00684161
Submitted on 2 Apr 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
The industrial organization of competition in local bus services

Philippe Gagnepain, Marc Ivaldi, Catherine Vibes

October 27, 2010

Acknowledgement: This chapter is drawn from a report commissioned by the Competition Commission and published at http://www.competitioncommission.org.uk/inquiries/ref2010/localbus/pdf/100929_Review_of_economic_literature.pdf. We would like to thank Robin Finer and his team for insights and remarks. The views expressed in the paper, and any remaining errors, are solely ours.
Introduction

In most countries, local transport services by bus, which are a major component of regional passenger transportation systems, are subject to the scrutiny of policy makers for at least two contextual reasons. First, while the passenger transport services have always been highly regulated, the public transportation policy is now experiencing deregulation and/or privatization in an industry where urban transport companies are heterogeneous in their ownership status, which can be public or private, as well as in the diversity of transport modes they offer (bus, train, underground, and tramway). Second, while the modal share of bus transport services has been declining for several decades in most developed economies, the growing environmental concern raises the calls for promoting urban mass transit (as opposed to private car).

This chapter is aimed at deepening our understanding of the functioning of competition in the local bus transportation industry and to evaluate its effectiveness. It provides an overview of the competitive constraints that are at work in the industry as discussed in the economic literature, and sketches empirical tests to check whether the intuitions provided by the economists are in line with the reality of the industry. To address these various issues, the first three sections of this text survey the economic literature on bus competition, emphasizing the case of UK which is used as a benchmark. We suggest that earlier contributions, proposed in the late eighties, (i.e., just after the deregulation of the industry) are very often based on unrealistic assumptions, mainly chosen because the authors lack of a sufficient perspective on the effects of deregulation. Hence, we focus on the most recent literature, which we attempt to survey as completely as possible. The objective is to draw the main conclusions or results which are shared by the analysts or researchers on how this economic activity functions. In the last section, we propose some methods to empirically test these main predictions of the economic literature.

The analysis is summarized as follows. Section 1 reviews what it is empirically known about the technological features and the economic performance of bus operators.¹
Section 2 focuses on the analysis of demand for urban transport services. Section 3 analyzes the competition in local bus service industry, leading to the conclusion that operators do not compete in prices but mainly in frequency. The last section concludes.

**Cost and production analysis**

The cost and production analysis addresses three main issues. First it is aimed at measuring economies of scale, economies of density and economies of scope which are key structural elements to describe in economic terms the technology behind an industry. Second, it should provide a measure of the level of technical efficiency to evaluate the performance of firms. Finally, it discusses the definitions for the output variables. In the literature, either supply indicators (e.g., vehicle-kilometres or seat-kilometres), demand-related output measures (e.g., passenger-kilometres or the number of passengers) or multidimensional output definitions are used. (See De Borger and Kerstens (2006) for a discussion on the choice of output measures.)

**Technological characteristics**

Let us first review some characteristics of the technology for bus transport services, such as returns to scale, economies of density and economies of scope. It is a common result that bus companies experience increasing returns to scale. (See Filippini and Prioni 2003, Farsi *et al.* 2006, and Farsi *et al.* 2007.) More specifically, it seems that smaller firms benefit from increasing returns to scale, as opposed to larger firms which exhibit constant or even decreasing returns. (See Viton 1997, Matas and Raymond 1998, and Kerstens, 1999, .) For the British bus industry, Cowie and Asenova (1999) estimate that small companies of fewer than 200 buses experience some economies of scale. They also find that the size of such returns varies with the company type whether it is public limited, private limited, or municipal. Sakano and Obeng (1995) find increasing returns to scale for the U.S. urban transit industry. Overall a significant number of empirical studies are in line with a U-shaped average cost function exhibiting increasing returns to scale for the smaller operators, which become constant and finally decreasing as companies’ size increases.
In most empirical studies, economies of density are frequently found regarding the bus companies’ technology. As already pointed out, the distinction between economies of density and economies of scale is very important in industries that provide their services over a network. In these cases, the firm size is more closely related to the size of the network than to the output provided over that network. For this reason it is important to distinguish cost changes that occur because of output changes only and cost changes that occur because of a proportional network and output change. Among studies which estimate that bus companies fail to operate at an efficient density are Farsi et al. (2006) and Filippini and Prioni (2003) on the Swiss market, Matas and Raymond (1998) for Spain and Shaw et al. (2005) for Taiwan. It appears that bus operators could obtain cost-saving benefits by extending their output scale.

Some articles have focused on the multi-modal side of the industry and have asked whether a bundling of operations from different urban transport modes (bus, train, metro for instance) is preferable to a separated configuration. They converge to the conclusion that economies of scope are significant in the industry, and that their results are in favour of integrated multi-modal operators. Farsi et al. (2007) conducted a study in Switzerland and found increasing returns to scale in almost all outputs. They consider that these returns, combined with cost complementarities, can be considered as a suggestive evidence for natural monopoly. Viton (1993) also finds positive economies of scope and concludes that together with the nature of economies of scale, they support the formation of larger multi-modal systems in the San Francisco Bay Area.

**Efficiency**

The recent literature on performance of operators of local bus services shows that there still exists a substantial level of inefficiency in this industry. However, huge differences exist over time and across countries. Cowie and Asenova (1999) find a high degree of inefficiency in the British bus industry which they interpret as an indicator of wasteful competition. However, Cowie (2002) estimates that the average efficiency has improved in the U.K., suggesting that mergers may have allowed existing group companies to operate closer to the optimal level of output. Heseltine and Silcock (1990) for the British operators find that the main total unit cost reduction was achieved by productivity
improvements. Working on a sample of Spanish cities, Garcia-Sanchez (2009) finds that a majority of municipalities are technically inefficient, mainly due to scale inefficiency. This is a similar result to Kerstens (1999) who indicates that inadequacies in scale are the major source of poor performance in her sample of French urban transport service operators. Some studies though are more optimistic in their measurements of efficiency, in particular in the UK. Viton (1997) finds that 80% of bus systems are efficient in the U.S. Wunsch (1996) who compares 178 European urban transport companies claims that two British firms, in the cities of Manchester and Sheffield, are among the first on his list in terms of technical performance. However, he takes into account only dominant bus companies and he admits that his result depends crucially on data quality. Most studies underline the dispersion in the efficiency measures they obtain within the same country or area. (See Kerstens 1999, De Borger and Kerstens, 2006, with the exception of Salas, 1998, who finds that, in Sweden, the levels of efficiency are very similar among companies.)

**Private / Public ownership**

Contrary to a common argument, there is substantial evidence in the literature that private bus companies do not operate more efficiently than public companies. Ownership type does not seem to be a crucial determinant in the firms’ performance, as shown in Odeck and Sunde (2001) and Garcia-Sanchez (2009) on the Norwegian market and Viton (1997) who shows that U.S. public and private systems share the same distribution of technical efficiency. Fazioli et al. (1993) found no relation between technical efficiency and ownership among a sample of Italian urban transit firms precisely because of the absence of effective competition for both public and private operators and strong regulation. Filippini and Prioni (2003) underline that the results in their study on a Swiss sample depend on the specification of output and network variables. However, if we can assert there is no strong evidence of a higher efficiency for private firms, some studies do estimate they perform better. Cowie and Asenova (1999) find privately owned firms are not more technically efficient, although they exhibit a considerable level of managerial efficiency. They find that values of increasing returns to scale for small companies not only vary with the ownership type (public/private) but also with the actual form of private ownership. Relevance of
ownership as a determinant for performance is also found in Kerstens (1996) and De Rus and Nombela (1997) on the French and Spanish market respectively. At this point, the literature is considered inconclusive regarding the impact of ownership type on efficiency.

**Subsidies**

There is some evidence that subsidies are associated with an increase of operating costs. In particular, Kerstens (1996) corroborates this assertion when analysing a sample of French urban transit companies. Sakano and Obeng (1995) on U.S. transit systems report that subsidies lead to excess use of labor relative to capital and excess use of fuel relative to capital and labor.

**Incentive contracts**

Several recent studies have revealed the positive effects of incentive contracts on technical efficiency. In Kerstens (1996), empirical findings confirm the importance of appropriate incentives in contracting for monopoly. Risk-sharing agreements seem to stimulate the performance of organizations. These results for French operators are confirmed by Gagnepain and Ivaldi (2002a) who develop a method which should help to clarify the choice of regulation in the urban transport industry. They conclude that cost-plus contracts are dominated by any type of second-best contract. These results are in line with those of Roy and Yvrande-Billon (2007) who find that operators under cost-plus contracts exhibit a higher level of technical inefficiency than operators under fixed-price agreements. De Borger and Kerstens (2006) survey other European studies which exhibit that high-powered incentive contracts improve efficiency.

**Competitive tendering**

The available evidence suggests that competitive tendering may improve performance. These results are exhibited by Hensher and Wallis (2005) who review the international successes and failures of competitive tendering from ten developed countries.

**Methodologies / Discrepancies**

It is important to bear in mind that all these performance analyses differ in several aspects. First, there exist several approaches to estimate efficiency on the basis of observed data. Efficiency, as measured by a deviation from the unobserved cost or production frontier, can be estimated by means of parametric and non-parametric methods aimed at determining the production or cost frontiers. On the one hand, parametric methods require the specification of a functional form for the frontier, a popular one being the flexible translog cost function. On the other hand, non-parametric approaches do not need to specify a functional form; they construct the frontier by enveloping the data on inputs and outputs by piecewise linear hyperplanes, as proposed by the extensively used data envelopment analysis (DEA) method. Both methodological strands have advantages and weaknesses, related to the presence (or not) of measurement errors or the requirement to specify functional forms. A detailed description and discussion of these frontier methodologies are presented in Lovell (1993) and Brons et al. (2005) respectively.

A second source of differences in the measurement of efficiency comes from the definition of the output variable. A significant number of studies conclude that operators’ performances differ substantially depending on the output specification considered. Supply indicators (e.g., vehicle-kilometres or seat-kilometres) or demand-related output measures (e.g., passenger-kilometres or the number of passengers) have been used.

A third crucial aspect in the model specification for measuring efficiency is that models should account for relevant measures of service and network characteristics. Bus-transit services have been recognized as very heterogeneous across countries and even cities. This is confirmed by Brons et al. (2005) and De Borger and Kerstens (2006) who find significant and consistent effects of the type of database, region and output measurement method. Fourth, some authors underline the need to decompose the measures of efficiency into two components (allocative and technical). For example, Viton (1997) suggests that the result of similar efficiency distributions between private and public firms might hide the fact that private systems would be more allocatively efficient. Also,
according to him, the distinction between managerial and organisational efficiency seems relevant in this industry, particularly in measuring the impact of ownership type on efficiency. This conclusion is confirmed by Cowie and Asenova (1999). (See also Gagnepain and Ivaldi, 2002b.)

**Further research**

Although the literature on measuring efficiency in the urban transport industry is extensive, some aspects still have to be investigated more thoroughly. An international comparison on the effects of deregulation and competition on efficiency would be of high interest. Also, only a few studies take into account the presence of other transport modes on the market. Indeed, the presence of economies of scope and the call for limiting private car traffic to the benefit of urban modes because of environmental policies make this multi-output aspect of the industry particularly relevant. Further analysis of the decomposition of efficiency into its several components to better understand the effects of ownership and deregulation on efficiency seems to be a next step in the research agenda.

**Demand analysis**

In this section, we review the values of bus demand elasticities found in the literature. We discuss different types of elasticities. First, we look at the own price elasticities. Note that the own price elasticity of the demand that a firm faces is always more elastic than the aggregate elasticity of market demand. This is because there are fewer substitutes for a product at the market level than at the firm level. An example would be the substitution between competing bus services on a market as opposed to substitution between different transport modes on this market. Second, we present the measures of cross-price elasticities of demand which evaluate the substitution patterns between competitors (transport modes or services). For example, the cross-price elasticity from bus to car tells us the percentage increase in car demand following a one percent increase is bus fare. In the bus-transit industry competition can come from other bus operators, as well as
other transport modes (train, metro, car...). Finally, we report the values of income and service quality elasticities found in the literature.

**Own price elasticities**

It is a common result in the literature that the demand for bus-transit is not elastic in the short run. Most studies on bus-transit own-price elasticities agree on a value of -0.4 and this result is summarized in four surveys on urban demand by Goodwin (1992), Oum et al. (1992), Dargay and Hanly (1999) and Balcombe et al. (2004).

These studies emphasize that authors now agree on the necessity to consider dynamic changes in these own-price elasticities. All studies allowing elasticities to vary over time, that is to say, allowing demand to adjust to changes in price in the long-run, have agreed that demand in the long run is more elastic than demand in the short run. The role of dynamics in urban transport demand is the objective of the survey conducted by Goodwin (1992) who estimates that long-run elasticities range between 1.5 to 3 times higher than short-run elasticities. He concludes that a time-dependent specification for the demand is preferred. In the literature (Goodwin 1992, Balcombe et al. 2004) the long-term response should be expected in a period of 5 to 20 years according to the authors. Only Matas (2004) on the Spanish market finds that 95 percent of the effects are realized within 3 years. As shown in Table 2.1, the values for long-term own-price demand elasticities vary from -0.4 to -1.3. However, values significantly greater than 1 are rare in the literature. Among the articles displaying the highest values are Romilly (2001), Dargay et al. (1999) and Gilbert and Jalilian (1991) on the British market. On the other hand, a study conducted by Deb and Filippini (2010) on the Indian market leads to relatively small values of long-run elasticities which the authors interpret as the effect of the low level of development in India and the fact that public transport is still a necessity there.

That long-term are higher than short-term elasticities has the following implications. First, the full behavioural response to fare changes cannot be properly identified by means of unlagged time-series models. Now demand models estimated on cross-section data can only reveal long-run price elasticities. Second, in this industry, the range of responses open to people is larger in the long run. Car ownership decisions require time to be implemented.
It is well known that this dynamic aspect of demand is an important consideration in implementing policy strategies.

Another important finding of the literature on own-price elasticities for bus-service demand is that the estimated measures vary with the type of ticket purchased by customers. The common result is that demand for a single ticket is more elastic than demand for a travel pass. Instead of building a price index to analyse the impact of a change in this price on demand, some authors have disaggregated these effects with respect to the different categories of tickets available to the customers. De Rus (1990) estimates fare disaggregated elasticities for bus-transit in Spanish cities and finds that data disaggregated by ticket fare provides a deeper understanding of demand responses. As he finds that price sensitivity decreases as we move from single tickets to the travel pass, he concludes that an aggregate approach fails to allow explicitly for shifts in demand between ticket types and that the role of cross-effects between ticket-types is key for the pricing policy. These results are in line with other studies on aggregate data, like Tegner and Holmberg (1998) on the Swedish market, and on micro data as in Hensher (1998) and Taplin et al. (1999). However, these last two analyses report smaller values for the elasticities. Matas (2004) in a more recent article with aggregate data on the Spanish market confirms these previous results and concludes that there is scope for a more efficient non-uniform pricing policy with positive effects on demand while minimising the negative effects on revenue.

**Cross-price elasticities / Substitution effects**

A change in fare for a transport mode can lead a customer to switch to another competitor, within the range of all available urban transport modes available (private car, train, bus, metro, or others). These substitution effects between travel modes are important when analysing competition and we present here the main literature findings on these measures.

The common result in the literature is that these substitution effects between modes are of a small magnitude in the short run. However, some authors consider that these findings, combined with higher long-run own-price elasticity for car and bus use, make modal shifts more feasible than often assumed (Goodwin, 1992). Hensher (1998) who distinguishes between fare classes finds that, in the Sydney metropolitan area, the largest
cross-elasticity between private car and train travel pass is 0.335 in the event of an increase in the price of car utilization. He also finds that there are more changes between modes for a given fare class than between fare classes within modes. The strongest cross-mode substitution for a given fare class (excluding car) occurs between train and bus single tickets with cross-elasticities of 0.067 and 0.057 for train-to-bus and bus-to-train respectively. Taplin et al. (1999) who aim at improving the methodology presented in Hensher (1998) estimate that the most significant differences observed between the two approaches are a large decrease in the elasticity of demand for car with respect to the price of a ticket for a single trip called Bus Single (from 0.066 to 0.018), and a large increase in Bus Single with respect to car cost (from 0.116 to 0.212). Matas (2004) looks at cross-price elasticities between ticket types and between transport modes. According to his results, bus users are sensitive to both bus and underground prices and quality, whereas underground users are only sensitive to underground characteristics. However, he also concludes that there is not enough information to understand the impact on modal shifts from car to public transport. Dargay and Hanly (1999) observe that the cross-elasticity between bus patronage and motoring costs appears to be negligible in the short run and about 0.3 to 0.4 in the long run. According to them there is some price substitution between bus and car use, although comparatively small. Balcombe et al. (2004) find that, in urban areas outside London, public transport use is sensitive to car costs but car use is much less dependent on public transport costs. Oum et al. (1996) in a study of the Dutch urban market estimate that the relative price of private car must rise significantly to induce a significant number of car drivers to switch to public transport modes.

*Trip purpose / Peak and off-peak demand*

Fare elasticity is different for different journey purposes. Trips made to go to work or to school are considered as peak demand, whereas trips for leisure or shopping are much more flexible in the time of the day and correspond to an off-peak demand. One would expect fare elasticity to be higher for off-peak demand than for peak demand where customers do not have much choice but to travel. In their review of the literature, Balcombe et al. (2004) observe that the mean off-peak elasticity for buses (precisely, -0.5) is at least twice the peak elasticity (i.e., -0.2). This is in line with the World Bank report by Oum et

**Income elasticities and car ownership effect**

Dargay and Hanly (1999) observe that, in the UK, the income elasticity for bus services, which includes car ownership effects, is negative in the long run. This is in line with the literature and suggests that bus transport is an inferior good. (See Bresson et al., 2003, Balcombe et al. 2004.) The negative long-run elasticity reflects the effect of income through its positive effect on car ownership and use, and the negative effect of the latter on bus patronage. They estimate that income elasticity ranges between -0.5 to -1 in the long run. However, as car ownership approaches saturation the income elasticity can be expected to become less negative. Romilly (2001) finds a positive value of 0.61 for his long-term income elasticity, suggesting that the economic growth has outweighed the inferior good aspect of the service. Matas (2004) also finds a positive value for the income elasticity (precisely, 0.15) in Spanish cities. He explains the difference with Dargay and Hanly (1999) by the higher population density of Spanish cities which makes them better suited to public transport use than to car use.

**Service Elasticities**

Regarding service elasticities, Matas (2004) estimates a service elasticity of 0.24, although he explains that, in aggregate studies, a very crude proxy for the quality of service is used, and it is difficult to give an adequate interpretation of the estimated elasticities. Quality is defined in different ways in different studies making uneasy the comparison of their values which are ranging between 0 and 1. De Rus (1990) finds a high coefficient of variation between the different cities. According to Deb and Filippini (2010) and as expected from the literature, service quality is the most significant policy variable as it has the largest impact on travel demand. Bresson et al. (2003) show that, in France and in the U.K., fare and service elasticities are of a similar magnitude (although opposite in sign), so that an increase in fares combined with an equivalent increase in service (vehicle kilometres) would have only marginal effects on patronage.
Methodologies / Discrepancies

Several approaches are used in the literature to compute reliable measures for urban transport demand elasticities. There is common agreement that variances in values for the different elasticities are influenced by several factors, both related to methodological aspects and to features of the industry. In particular Nijkamp and Pepping (1998) have carried out a comparative analysis of elasticity values of transport demand resulting from twelve studies in various countries. Their analysis indicates that the difference between aggregate (macro-) and disaggregate (micro-) models, as well as with other assumptions, explain the variance in the values of elasticities across studies. They also find that the country involved, the number of competitive modes, and the type of data collected are important factors in accounting for the level of elasticities. These conclusions confirm the findings of Oum et al. (1992) who survey the elements that impact the estimation of demand elasticities in different studies. Oum et al. emphasize the need to take into account intermodal competition because, otherwise, own-price and cross-price elasticities are biased given that they ignore some of the competing services. They also underline that different functional forms can result in widely different elasticity estimates, even with the same set of data. Note that models also differ with the choice of the definition of the dependent variable (whether one considers journeys or passengers-kilometres) and the way fares are aggregated into a price index. They observe that results differ according to the area or country under analysis, which have their own features (in particular for their urban-transit services). This is why they highlight the fact that disaggregated data would lead to a wide range of elasticities as they would reflect unique market conditions. Dargay and Hanly (2002) find a considerable variation in the fare elasticity across counties, ranging from 0 to -3 in the long run. Bresson et al. (2003) in their comparative study between France and the U.K. confirm the relevance of taking into account countries’ heterogeneity. The study by Dargay and Hanly (1999) corroborates the findings of Oum et al. (1992) and Nijkamp and Pepping (1998). First they find a large variance of elasticities across counties in the U.K.; second, they conclude that estimated elasticities from different studies are not directly comparable. More precisely, they assess it is inappropriate to apply the value of an
estimated elasticity for different circumstances or to average the values of elasticities from different studies.

**Further research**

The preceding review of the literature on urban transport demand highlights some areas for improvement in the methodologies adopted so far. First, models for disaggregated data have rarely been estimated and they would constitute a considerable enhancement in urban demand studies. They would allow us to capture the specific effects of the markets under scrutiny, such as different ticket fares, trip purposes, and customer categories. An aggregate elasticity hides these specific effects. Second, more structure could be applied to the models and the interaction between supply and demand could be taken into consideration. Third, the literature suggests that a comprehensive representation of the market is important as we observe significant differences in characteristics across cities. Competition from other modes should be taken into account to avoid bias in the measures of elasticities. Fourth, functional forms have to be chosen carefully as they can lead to very different results, even applied to a same dataset. Econometric testing of different model alternatives would be a useful part of the research agenda.

**Competition analysis**

**Entry**

As a general rule, a firm enters the market only if it can earn positive profits. When entry happens on a significant scale, it is expected that the incumbent reacts. In the early deregulation period, the literature focusing on bus competition suggested that entry may be a relevant issue and has shed light on several cases of entry in local markets. Entry usually occurs on the periphery of the incumbent’s main market area, particularly if the incumbent has a local reputation. Some smaller operators have attempted to enter on a small scale hoping not to invoke a response from the incumbent firm. The literature suggests however that entry strategies have been unsuccessful in most cases. (See Preston, 1988, for an early
To explain why entry was unsuccessful in the early deregulation period and why it was scarce in the years following deregulation, an important argument is that the industry is not perfectly contestable. As an indication that the industry is not contestable, we list as a first step the usual conditions which guarantee that a market is perfectly contestable. As a second step, we discuss why these conditions seem not to be met in the bus transportation industry.

According to Shepherd (1984), Baumol (1987), and Banister (1997), a perfectly contestable market requires the following conditions:

- Entry is free;
- Entry is perfectly reversible, i.e., sunk costs are zero;
- The incumbent and the entrant have access to the same technology;
- The incumbent and the entrant have equal access to all customers in the market, i.e., consumers are not loyal to the incumbent’s products; the services of the incumbent and the entrant can be easily accessible (for instance bus terminals can be used by all operators);
- There is an active second hand market for capital assets (e.g., the entrant has access to “cheap” buses for its rolling stock);
- The regulator imposes time lags to prevent sudden changes in prices or withdrawal of services by the incumbent firm, e.g., “hit and run” strategies, where the entrant enters the market over a short period and enjoys high prices, can be implemented.

In a contestable market, any attempt by incumbent firms to earn excessive profits would be unsuccessful. Furthermore, even if there is just one firm offering the service, this firm would be engaged in average cost pricing and have zero profits. If positive profits were obtained, competitors would enter the market and undercut the incumbent’s prices and profits. Hence, the important idea is that the mere threat of entry forces the incumbent not to behave as a monopoly despite the intrinsic properties of the market which enable it to do so.

The economic literature is unanimous in stating that the local bus transportation industry is not perfectly contestable. In the very first years of deregulation, Preston (1988),

analysis. Note that the literature does not provide any further evidence of successful entry in the 90s or the 00s.)
Button (1988), Beesley (1990), and Evans (1991) suggest that many factors prevent the markets from being contestable:

- **Existence of barriers to entry:** The access to bus stations and the use of travel cards have acted as barriers to entry; the incumbent may have more convenient terminal positions; entrants may not obtain access to bus stations; information points may be manned solely by the incumbent firm’s staff and entrant firms may be located at the least attractive stands in the bus station; other practices include the blocking of a rival’s bus, occupying a stand or using couriers to persuade customers to use one company’s buses in preference to another. Barriers to entry may have been underestimated at the moment of deregulating the market.

- **Existence of sunk costs:** Trained staff (managerial, administrative and platform) is costly. An entrant finds it difficult to hit and run if its employees are not highly qualified.

- **Practices which raise rivals’ costs:** Operators may withhold surplus buses from the second hand market, hoping that the price of old buses would increase as their availability decreased, hence making entry to the local market more difficult.

- **Economies of experience, economies of scale, economies of density, and economies of scope:** The incumbent may have a larger network than the entrant and may therefore be able to offer more attractive area-wide tickets than the entrants; the incumbent may be better known.

- **Incumbents can reduce prices very quickly (usually within 24 hours).**

  During the 90s, these initial intuitions are confirmed. Evans (1990 and 1991) insists on the fact that the incumbents can change their prices immediately in response to entry since operators are allowed to change fares without notice. As a result, incumbents can enjoy super-normal profits on high density routes. Moreover, the “experience” input is essential to explain the tactical advantage of the incumbent firm, given that it is usually better informed about different aspects of providing the service. Beesley (1990) claims that barriers to entry are numerous: For instance, garage locations and other property rights play a key role since they directly affect the likelihood that local markets can be opened to competition. The law may itself impede entry: For instance, entrants are required to remain at least 6 weeks in the market; sub-contracting to drivers is restricted.
Banister (1997) contributes to this view of the industry when he states that the characteristics of the industry and the strategic actions of the incumbent both impede the local transport market from being contestable. In addition to the previous factors, Banister proposes the following characteristics:

- The need to replace the ageing bus fleet, which requires greater capital investment than the smaller companies are able to obtain or willing to risk;
- The fear of competitive disadvantage of the smaller operators against the larger operators is significant in the bus industry;
- Large and small companies do not have access to finance on equal terms. Incumbents tend to have weaker risks of bankruptcy than entrants because they have a larger size and have a bigger purse. The incumbent may own routes elsewhere which earn high profits which can be used to cross subsidize more less competitive routes;

At the same time, Banister sheds light on the possible actions to be taken by the incumbent to reduce the arrival of entrants; these actions are:

- Build up consumer loyalty;
- Establishing a reputation for toughness by maintaining a presence in the market;
- Reorganize the network so that economies of scale and density can be obtained;
- Maintain ownership of fixed assets such as terminal, booking and maintenance facilities.

Banister thus concludes unambiguously that “the theory of contestable markets does not apply to the bus industry. In 1985 it may have been attractive to accept the contestability arguments, but this does not seem to be true anymore ten years after, since the size of operations seems important. The role of the small operators is reduced to competing through the tendering process for the socially necessary services.”

After 2000, the initial propositions listed above, on why the industry is not contestable, are corroborated and new claims are made on this issue. First, De Borger and Kerstens (2006) suggest that the rolling stock capital of entering firms has the characteristics of a sunk cost. More importantly, the incumbent’s strategic actions impede entry:

- Incumbents can easily cut prices and adjust schedules;
• The incumbent operates the fixed facilities (a central bus station for instance) available that are crucial to exploit network economies (interconnections between different lines or sets of lines), given that the demand structure is characterized by complementarities between lines.

Second, Langridge and Sealey (2000) emphasize the idea of the economies of experience enjoyed by the incumbent. They note for instance that the confederation of Passenger Transport in the UK (the major lobbyist for bus operators, see http://www.cpt-uk.org/) believes that the incumbent operator always has an advantage over the entrant through knowledge and experience, resources (staff), infrastructure, and reputation.

New strategic behaviours are emphasized as well. Some of them are related to the idea of combining competitive services and subsidized concessions allocated to operators though competitive tendering. In particular, Langridge and Sealey (2000) note that entrants could minimize barriers related to lower knowledge and experience by entering from a contiguous market in which they had already gained some knowledge and experience and/or entering a local bus market on a small scale, which could be achieved by obtaining contracts with the local authority. At the same time, many incumbents are eager to enter into the new quality partnerships with local authorities, even if this entails supporting high costs of investment in new vehicles and related infrastructure. This suggests that they are looking for long term partnerships through the creation of local monopolies.

Finally, as suggested by Van der Veer (2002), under entry threats, the incumbent may run more buses and increase the frequency of the service (compared to a situation where it is protected from entry) to avoid leaving profitable gaps. Wang and Yang (2005) corroborate these findings; they suggest that deterrence through an increase of the service level is a dominant strategy for an incumbent under various market conditions, which in turn explains the high levels of service in many industries. Accommodation occurs mostly on routes where demand is high. Blockaded entry occurs on routes where demand is low.

**Competition in price or frequency**

Early theoretical models on bus competition have usually been based on strong assumptions which were in most cases unrealistic:

• All operators face the same costs;
• All operators and passengers have complete information about services and fares;
• Operators have information about demand;
• Each passenger has a preferred departure time and is indifferent between immediate backwards and forward rescheduling;
• Traffic conditions are such that journey times are the same throughout the day;
• Departure times and fares of other operators are fixed.

(See Evans 1987, and Preston 1988, for a survey.)

Moreover, it has been suggested that service quality matters and is therefore a key factor in bus competition. (See Dodgson et al. 1992 and 1993, Dodgson and Katsoulacos 1988, Bly and Oldfield 1986, and Glaister 1985 and 1986.) In particular, minibuses have been considered as relevant actors in theoretical frameworks with quality differentiation, where competition can be implemented on a horizontal perspective where firms compete in fixed time schedules and prices. Regular buses were thought as cheap and slow services, while minibuses were associated with lower travel time and higher prices.

These different assumptions have been, to various degrees, criticized later on. The most important criticisms have been related to the assumptions of quality differences and price competition. Preston (1988) suggests that consumers have difficulties in perceiving quality differences. Moreover consumers’ loyalty to a particular firm seems to be unrealistic: Users usually board the first bus that arrives. A model’s outcome of two firms offering distinct qualities of service and charging different fares has not been as common as might be expected. Such a model of competition would probably be more relevant in explaining inter-modal competition.

Thus, it seems to have been accepted that competition has tended to take the form of service wars with fares matching. Passengers board the first bus that arrives, hence making frequency the key factor for competition. Competition in fares has been mainly restricted to branded ticketing such as system passes, return ticketing, multi-rider tickets or discount vouchers; branded ticketing is thus seen as a tool for operators to increase the consumer’s incentives to be loyal to one specific company.- it is an attempt by operators to develop strategic barriers to entry. (See Fernández and Muñoz, 2007.)

Later on, many authors, such as Van Reeven and Janssen (2006) and Wang and Yang (2005) have confirmed these early intuitions. Price competition (and therefore price
reduction) is not particularly prevalent in the bus industry.\textsuperscript{vii} Operators have limited scope for meaningful product differentiation that could make consumers loyal.

However, on long distance services such as intercity bus services, consumers’ loyalty and price competition (through higher services quality) are more relevant. In this case, quality matters, and ticket prices constitute an important fraction of the generalized price paid by consumers. Hence, product differentiation on long-distance routes makes entrants resistant to pricing and scheduling responses of incumbent operators. Scheduling competition is more stable in this case.

**Random schedules**

The previous section suggests that competition mostly takes the form of frequency wars. Analysts then go a step further when they explain that the arrival time of a bus at a stop is random.

Oldale (1998), Ellis and Silva (1998), Van Reeven and Janssen (2006) and Gomez-Lobo (2007) all agree on the fact that the incentives for price competition are smaller than what was expected, even if more than one operator is present on a local transport market. Two main reasons explain this result: First, users do not particularly care for quality difference, and second, they incur a cost if they want to shop around for the lowest priced bus. Contrary to Evans (1987) which assumes that operators’ services are scheduled, these authors consider some degree of uncertainty surrounding arrival times at bus stops. In their model, users arrive at a stop and will wait for the arrival of the next bus; an important assumption is that the distribution of passengers across time is uniform, i.e., there are no masses of passengers clustered around departure points. The optimal reaction of the bus operators consists then of randomizing arrival schedules at the bus stop, and setting the highest possible prices.

Given that consumers do not differentiate one bus company from another, random frequencies have to be expected for the following reasons: Some buses may bunch together or some may be alone at a given position in time and space. In the first case, each operator has an incentive to drive just in front of the others. Thus, bunching cannot be an equilibrium. A profile where each bus is alone in a position cannot be an equilibrium either,
since buses have an incentive to fall back and drive just in front of the next bus that is following behind. These techniques are known as head running and leapfrogging. Hence, randomizing the arrival at a bus stop is the best strategy for each operator competing on the same route, and this forces the rivals to guess the arrival time of their competitors. A striking example is the case of Manchester in the UK, where the first two years of deregulation were characterized by services changing between 1500 to 2000 times annually. Bus companies cannot credibly provide timetable information. In these conditions, competition does not guarantee low prices.

A research agenda

The economic literature proposes a number of arguments as to why bus competition might be limited in liberalized industries. Several reasons could explain such a situation: First, the technology used in the industry favours large and experienced operators and therefore impedes the entry of new competitors on an equivalent scale. Second, the fact that the transportation service occurs on short distances restricts the incentives of the consumers to look for the cheapest operator and/or the company offering the highest quality standards. Price competition is therefore likely to be very limited, even on routes where more than one operator is present. Note that, where there is no regulation providing incentives to bus companies to comply with the time schedules, there is no guarantee of a proper coordination of consumers at bus stops, which again limits the scope for competition.

As pointed out above, most of the contributions drawn from the economic literature are theoretically derived, although based on experts’ knowledge and experience. Although these theoretical arguments are intuitive and convincing, they often wait to be empirically validated. There are potential avenues of investigation.

First, one should test the contestability of the market. It is well known that, if a technology exhibits increasing returns to scale, then the associated industry is highly concentrated or is operated by a single firm. In this case, the presence of too many production units prevents the efficient size of the industry being reached, which could be socially costly. Without entering into the details of this theory, this result invites us to
evaluate the level of economies of scale and scope, which is usually performed by means of the estimation of cost functions. To do so one can use data at the depot or firm level. Now, the question of the adequate level of disaggregation to measure output in a network industry is still open and subject to research. However, if estimating firms’ cost functions is crucial to characterize the economic fundamentals of the bus industry, recall that such an exercise is not immune from other constraints that the firms are facing, like the regulatory conditions. For instance, the local regulators may impose different quality targets, or the driving conditions may vary from one urban network to another, and this may explain cost differences across local areas.

Second, another set of evidence supporting the non contestability of the bus industry can be achieved though the detection of predatory pricing: Following Motta (2004), predatory pricing implies that the incumbent sets low prices for a period and sacrifices short-run profits, so that the entrant believes that positive profits cannot be obtained. When the entrant leaves the market, the incumbent then increases prices and reaches high profits again, which in the long run outweigh possible losses incurred by foreclosing entry. Note that observing that entry occurs is not enough to conclude that the market is competitive or that there are no predatory practices. To properly detect predation, prices should be compared to marginal and average costs. Following Motta, a test of predation could be implemented as follows. First, from the estimated cost function, we can evaluate total and marginal costs. Second, actual prices must be compare to these estimated costs: i) If the price is above total average costs, then the presumption is that the firms are not taking predatory actions; ii) if the price is below total average costs but above marginal costs, then predation should not be presumed, but the burden of proof is on the side of the competition authority; iii) If the price is below marginal costs, then there is a case for predation. Again note however that these tests should not be applied without taking into account the regulatory and competition constraints. Indeed regulation of prices and services or competition from other transport modes could clearly affect the pricing strategies of bus companies.

Third, one should test the theoretical prediction that, companies do not compete on prices at the route level in the short run. There are at least two ways to test this assertion. There is a direct approach that consists of estimating a structural model of the industry that
comprises a demand function and a pricing equation. This approach could be implemented using models specified along the line of the econometrics of differentiated products markets. (See Davis and Garces, 2010, for a presentation of these models.) There is also an indirect approach that it is easier to carry out. It is indirect in the sense that it tests a necessary condition not a sufficient condition, namely that the number of firms on the market has no effect on the price level. This approach relies on the structure-conduct-performance paradigm which states that the structure of a market determines the operators’ pricing conduct and therefore their profitability. The ability to obtain significant profits is inversely related to the number of firms and/or their market share, and thus is positively correlated with concentration. If operators do not compete on price, as suggested by the economic literature, a non-significant long-lasting relationship between the price and the variable measuring the degree of presence of firms on the market should be obtained. Note that one should similarly test for the impact of the market structure on frequencies.

The research agenda is thus particularly rich. It is also urgent as the effectiveness of competition in the bus industry becomes a crucial issue in many countries over the world.

References on cost and production analysis


References on demand analysis


Deb, K. and M. Filippini, 2010, “Public bus transport demand elasticities in India”, Quaderni della facoltà di Scienze economiche dell'Università di Lugano, number 1002.


Deb, K. and M. Filippini, 2010, “Public bus transport demand elasticities in India”, Quaderni della facoltà di Scienze economiche dell'Università di Lugano, number 1002.


References on competition analysis


---

i Note that De Borger and Kerstens (2006), Brons et al. (2005) and Berechman (1993) have surveyed the literature on technical efficiency for bus transit.

ii Note moreover that, during this period, several factors have favoured entry; these factors are: The management of the entrant firm has personal knowledge of the area chosen for entry; or the entrant may have
hired former employees of the incumbent firm. Beesley (1990) notes that the population density and the incumbent’s initial market power are other factors which influence positively the likelihood of entry.


Highly skilled employees are so important in the production process that it is not uncommon to observe bus operators attempting to recruit a rival’s staff by offering higher wages and better work conditions.

They also shed light on the fact that, if the incumbent is unsuccessful in the tendering process, it may attempt to provide subsequently a commercial service in order to force the withdrawal of the rival of the tendered service.

See also Nash (1985), for a discussion of Glaister’s assumptions. A more recent contribution on differences in service quality is Yang et al. (2001).

Note that, currently, there are a number of low cost/‘no frills’ bus companies in towns across the UK that compete with higher quality offerings by the larger operators. (e.g. Whippet bus in Cambridge).

Recently, price competition models have been proposed by various authors. See for instance Zhou et al. (2005). Their model is however more relevant to describe bus operators’ habits in developing countries such as China and other Asian countries or modernized cities with high-density population such as Hong Kong and Singapore.

Other older “bad habits” of bus operators are discussed in Forster and Golay (1986). They entail “hanging back” (the buses go slowly so as to pick up as much traffic as possible), “missing out a bus stop” (if the driver decides that there are to few passengers to stop for), “turning” (an nearly empty bus turns around before the end of the route and go back in the opposite direction), or “overtaking”.