Demand Estimation with Selection Bias due to Exit/Takeovers: A Dynamic Game Approach with an Application to the US Railroad Industry

Daniel COUBLUCQ

Toulouse School of Economics

June 27, 2011

ABSTRACT

This paper is motivated by the US freight railroad industry, which is characterized by a major restructuring over the last 30 years. In particular, the number of active firms decreased from 26 in 1978 to seven in 2006 due to several takeover waves. The empirical focus concerns the estimation of a structural demand model for the US railroad industry. Then, the demand estimates are used to compute the evolution of the mark-ups, the quality of the freight services provided, and the consumer surplus. The restructuring of this industry involves significant exit and takeovers. This implies that the data is characterized by an attrition issue, which generates a selection problem. A focus is to provide an estimation algorithm which takes explicitly into account this attrition issue. I find that the algorithm produces more plausible estimates of demand coefficients compared to standard estimation procedures. Moreover, using the model, I recover the evolution of marginal costs, mark-ups, and consumer surplus over time. I find that the takeover waves have led to efficiency gains by decreasing the marginal costs, and this was translated into lower prices and an increase in the consumer surplus. Finally, the takeovers have led to a reallocation of assets from the less efficient firms to the most efficient firms, which improved the quality of the freight services provided.

Keywords: selection bias, panel data, demand model, merger/takeover analysis, railroad industry.
JEL Classification: C23, C51, L10, L41, L92.

1 I thank my advisor Marc Ivaldi for his encouragement and suggestions during the process of this work, and Gerard McCullough for providing the data and his guidance regarding the data construction. I also thank Mehtap Akgüç, Sofronis Clerides, Peter Davis, Yinghua He, Cristian Huse, Thierry Magnac, the participants of the ZEW conference “Quantitative Assessments in Competition Policy,” DIW conference “End of Year Summit 2010,” “ENTER Jamboree 2011” at Tilburg University, “8th CEPR Applied IO School” at Tel-Aviv University, and the internal “TSE Student Workshop 2011” for useful comments. All remaining errors are mine.

2 E-mail: daniel.coublucq@tse-fr.eu; Toulouse School of Economics, 21 Allées de Brienne, 31000 Toulouse, France.
1 Introduction

The US railroad industry is characterized by a rather “light” regulation. This regulatory freedom came from the Staggers Act, which deregulated the US railroads in 1980. In particular, this deregulation process came with several takeover waves that have led to a concentrated industry today. Indeed, there were 26 firms in 1978, while there are seven firms today.

The economic motivation of this paper is to examine the impact of past takeovers on the consumer welfare. In this framework, the consumers could be farmers, manufacturers, and so forth, who ship the goods to another place. In particular, the paper asks the following question: how has the concentration process impacted the consumer welfare over time? As mentioned in Whinston (2007), a retrospective analysis is a critical step for understanding past takeovers and improving future enforcement practice. This is related to the current debate regarding the state of competition in the US railroad industry, in which there are concerns about the potential market power of the firms and its impact on the consumers (see the GAO Report (2006)). In this vein, this paper analyzes the consumer welfare and its evolution by taking into account the past takeover waves, hence the concentration process, between 1980 and 2006 in the US railroad industry. While doing this, three important dimensions of consumer welfare are highlighted.

The first one is related to the pricing of the railroad companies and its implications for the consumers. According to the trade-off suggested by Williamson (1968), if the consumers have benefited from this concentration process, one might conclude that the railroad firms have passed the productivity gains due to takeovers to the consumers through a decrease in the final prices. Otherwise, it might be that the railroad firms have used the concentration mainly to decrease the competitive pressure and increase their mark-ups.

Secondly, although the main variables of interest is the price of freight services and its evolution over time with its impact on the consumer surplus, it is also important to incorporate the investment of the firms in the network since it impacts the quality of the network infrastructure and might be altered
by the takeover waves. To this end, I construct a variable that represents the quality of the network infrastructure, which impacts the quality of the freight services provided and hence enters into the consumer welfare function.

The last, but not the least, aspect of the paper during the analysis of the consumer welfare is to consider the takeovers as a way to reallocate assets from firms with low asset exploitation abilities to firms with high exploitation abilities. In this sense, the dataset on the US railroad industry is particularly suited for this purpose since this industry is characterized by an important attrition due to exit and takeovers (from 26 firms in 1978, only seven firms remain in 2006). Moreover, this redeployment of assets might lead to a selection bias in the sample since the least efficient firms leave the industry and the most efficient ones stay. This is an essential issue that this paper takes into account. The framework developed in this paper focuses on selection issue due only to exit and not to entry. This is justified by the data on the US railroad industry, which is characterized only by exit. However, the framework developed in this paper can easily be adapted to any selection issue due to entry, as well.

In particular, I provide an estimation algorithm to take into account this selection issue using a dynamic model, which allows for the equilibrium force of exit and from which a structural selection equation is derived. The selection equation yields a correction term, which is then plugged into the structural demand model in order to correct for selection bias. Intuitively, this can be seen like a standard two-step Heckman procedure, where the correction term is derived from an explicit dynamic model of exit behavior. While doing this, I build on the dynamic models of Aguirregabiria and Mira (2007), Bajari, Benckard, and Levin (2007), Berry, Ostrovsky, and Pakes (2007), Doraszelski and Satterthwaite (2010), Ericson and Pakes (1995), and on the literature on selection bias (Wooldridge 2002). The interested reader can look at Andrade et al. (2001), Andrade and Stafford (2004), Jovanovic and Rousseau (2002), Maksimovic and Phillips (2001), and Olley and Pakes (1996). This literature shows that mergers/takeovers are a way to reallocate assets and achieve productivity gains since the least efficient firms leave the industry and sell their assets to the most efficient firms and thereby put the resources to their best use.

For instance, it is reasonable to expect that any entry of new products/firms might lead to a selection issue since they are likely to be more efficient than the existing products/firms.
(1995), Semykina and Wooldridge (2005)). The idea of using a structural model to account for selection is also an important contribution of Olley and Pakes (1996) in the framework of production function. Another related paper is Mazzeo (2002), where a structural model of entry is used to deal with endogeneity of the market structure variables, i.e. the number of product-type competitors, in a reduced form price regression.

Furthermore, I also emphasize the connection between the demand model and the dynamic exit model. In particular, the variables that enter the demand model influence the spot profit function and thus the value function. In other words, all the variables that are in the demand model must also appear in the dynamic model. To deal with this issue, I carefully define the observed and unobserved state variables in order to ensure that the demand model is fully coherent with the dynamic model and this leads to an iterative algorithm. Lastly, I also explicitly discuss the link between the exogenous process of the unobserved state variable and selection bias. Indeed, the unobserved state variable plays an important role and it is interpreted as the unobserved efficiency/ability of a firm to provide freight services of good quality.

Moreover, using the parameter estimates from the demand model, I am able to recover the evolution of the marginal costs, mark-ups, and consumer surplus over time. I find that the takeovers led to important efficiencies with an important decrease of the marginal costs during the period 1980-2004. This decrease in the marginal costs was translated into a decrease in the freight prices for the same period. For the period 2004-2006, which is characterized by a price increase, I argue that it might be due to an increase in the industry marginal cost. This implies that the increase in freight price was rather a consequence of an increase in the marginal costs instead of an abuse of market power by the railroad firms. This is relevant for the current debate regarding the potential market power of the railroad firms in the US (see GAO report, 2006).

Furthermore, during the takeover waves, the consumer welfare has increased, which can be justified by two elements. The first element is the cost efficiency gains. Indeed, this industry is
characterized by important returns to density (see Berndt et al. (1992, 1993a, 1993b), Ivaldi and McCullough (2001, 2008)). In addition to the gains from the consolidation of the traffic, Dennis et al. (2010) mention other efficiency gains of takeovers in the sense of more efficient use of equipment by rerouting traffic to avoid congestion on routes, reduction of ton-miles by choosing the shortest routes, routes dedicated to specialized services, and so forth. This led to a decrease in marginal costs, which resulted in a decrease in final prices. The second element is the efficiency gains in terms of the provision of the freight services, which is the ability of firms to provide freight services of good quality, due to the reallocation of assets from the less efficient firms to the more efficient firms. Indeed, between 1980 and 2006, the takeover waves led to an increase of efficiency and it had a significant impact on the consumer welfare. What is more, the results of the paper indicate that the reallocation of assets aspect seems to impact the consumer welfare more than the price aspect. This result is relevant in practice for merger investigation by competition authorities, where the price aspect is often the main focus. In other words, this paper shows that the reallocation aspect can be as important as the price aspect for the analysis of the concentration in the US railroad industry.

Finally, the evolution of mark-ups over time shows that the concentration process led to an increase in the mark-ups during the period 1980-1999. However, since 1999, the rise of the mark-ups has stopped. This coincides with the end of the concentration process. For example, the Surface Transportation Board, the US rail regulator, did not file the proposed mergers between two railroad firms, Canadian National and Burlington-Northern, in 1999.

The remainder of the paper is organized as follows: Section 2 describes the US rail freight industry; Section 3 states the problem of selection bias by presenting a demand model and describes the dynamic exit model that allows to derive a correction term for the selection issue, which is then used in the estimation; Section 4 describes the different estimation procedures depending on the process of the unobserved state variable; Section 5 presents the estimation results; Section 6 concludes. A detailed description of the data is available in the appendix.
2 Overview of the US railroad industry

The US railroad industry is composed of Regional and Class 1 railroads. The dataset covers only the Class 1 railroads (operating revenue in excess of 346.8 million US dollars in 2006), which account for 67% of industry’s mileage, 90% of its employees, and 93% of its freight revenue. Figure 1 illustrates the network configuration of the industry, where Class 1 railroads operate in many different states. The structure of the US rail freight industry is characterized by the integration of the network and the provision of freight services: the US freight railroads operate on tracks they own and maintain.

![Figure 1. Network configuration of the US rail freight railroads in 2009](http://www.cn.ca/en/cn-and-class-1-railroads-flash.htm)

The industry is characterized by a rather “light” regulation. This regulatory freedom came from the Staggers Act, which deregulated US railroads in 1980. The Staggers Act gave the railroads the possibility to adjust their rates and capital structures. This deregulation process came with several takeover waves that led to a concentrated industry today. Namely, there were 26 firms in 1980, while there are only seven firms today (see Figure 2 and Figure 3, Table 5 and Table 6 in the data appendix, and Waters (2007)).

\[5\]

Namely, they are: Burlington Northern and Santa Fe Railway Company (BNSF), Kansas City Southern Railway Company (KCS), Union Pacific Railroad (UP), Soo Line Railroad Company (SOO) which represents the U.S. operations of the “Canadian Pacific” railways company, CSX Transportation Inc. (CSX), Norfolk Southern Combined Railroad Subsidiaries (NS), Grand Trunk Corporation (GTC) which represents the U.S. operations of the “Canadian National” railways company. Source: Surface Transportation Board (STB).
Figure 2. Railroad firms in the Western area
Railroads (class 1 and regional railroads) account for 41% of freight ton-miles, more than any other mode of transportation. Figure 4 illustrates the increasing importance of Class 1 railroads in the US national freight market, where the market share of the Class 1 US railroad firms increased from 20% in 1980 to 38% in 2006. The US national freight market is defined as the freight provided by air, truck, railroad, water, and pipeline, in the ton-miles unit.

**Figure 4. Evolution of the Class 1 railroads market shares in the US freight national market**
Coal is the most important commodity carried by US railroads. In 2007, coal accounted for 44% of rail tonnage and 21% of rail revenue. Coal accounts for around half of the US electricity generation and railroads handle more than two-third of US coal shipments. Other major commodities carried by rail include chemicals (ethanol, plastic resins, fertilizers); agricultural products such as grain; non-metallic minerals such as phosphate rock, sand, and crushed stone; food products; steel; forest products (e.g. paper); motor vehicles and motor vehicle parts; waste and scrap materials.

Several papers have studied the US railroad industry from a cost-side perspective. The cost function analyses of the 1990s —Friedlaender (1992), Berndt et al. (1993a), Berndt et al. (1993b)— show that the economies of density are significant and substantially greater than one. This finding is confirmed by Ivaldi and McCullough (2001, 2008) in a different framework in which they consider three different types of freight services; namely, bulk, intermodal, and general freights, measured in car-miles instead of ton-miles. They find economies of density equal to 1.90. This seems to be a pervasive technological characteristic of the US railroad industry, which yields a natural tendency for firms to be concentrated.

Another interesting feature of this industry is the capital adjustment since the deregulation in 1980 (Berndt et al. (1993b)). Indeed, the Staggers Act provided the railroads with considerable potential to rationalize their capital structure. Specifically, Berndt et al. (1993b) look at the extent of capital adjustment after the Staggers Act. They find that cost savings from increments in ways and structures capital are not enough to justify the observed levels of way and structures capital. This implies an excessive capacity in the industry. As mentioned by Berndt et al. (1993b), as the regulatory freedom allows the firms to adjust their capital, the overcapitalization of the railroad firms seems puzzling. Berndt et al. (1993a) mention the possibility that cost savings from increments in way and

---

6 The returns to density are defined as $S_y = (\partial \ln VC/\partial \ln y)^{-1}$, where $VC$ represents the variable cost and $y$ represents the output in ton-miles. In the paper mentioned, the returns to density (in terms of mean-value for each firm) have a range from a minimum of 1.477 to a maximum of 4.274.

7 Before the Staggers Act, the industry was characterized by an excessive route network and was unprofitable.
structures capital may not fully reflect the benefit of the investment. In particular, it is the case if service quality enters the demand function and if it depends on the amount of way and structures capital. Then the shadow value of capital (which considers only a cost-perspective) underestimates the actual benefits of investment. In addition to reducing costs, the way and structures investment may enhance demand by allowing higher speed and better service thanks to the high quality rail. This can be used to justify the inclusion of network quality in the demand model (see the data appendix for the construction of this variable).

3 A dynamic model to correct for selection bias

As mentioned in Olley and Pakes (1996) and in Wooldridge (2002), when the panel is unbalanced, it is important to account for the nonrandom nature of the sample and this requires a formal description of why the panel may be unbalanced. This section presents a dynamic model of firm behavior which allows for the equilibrium force of exit in order to derive a correction term that will allow estimating consistently the parameters of the demand model.

3.1 The demand framework: selection issue

I first state the problem of selection bias by presenting a demand model that will be used to compute the consumer welfare for each time period between 1980 and 2006.

Following Berry (1994), I group different railroad firms into two groups and exclusive sets, \( g = 0,1 \), where \( g = 0 \) denotes the outside option, which is the freight provided by air, truck, water, and pipeline, and \( g = 1 \) denotes the group containing the railroad firms. The utility of a consumer \( i \) from choosing the railroad firm \( j \) is:

\[
     u_{i,j,t} = \delta_{j,t} + \zeta_{g,t} + (1 - \sigma_g) \epsilon_{i,j,t},
\]


where \( \delta_{j,t} \) is the mean-utility of choosing railroad \( j \) at time \( t \) and \( \epsilon_{i,j,t} \) is identically and independently distributed extreme value.\(^8\) The variable \( \zeta_{g,j} \) is common to all firms in group \( g \) and follows a Cardell (1997) distribution \( C(\sigma) \), with \( \sigma \in (0;1) \). The parameter \( \sigma \) represents the within group correlation of all the alternatives in the group \( g=1 \).

Regarding the mean-utility, \( \delta_{j,t} = x_{j,t} \beta + \theta k_{j,t} - \alpha p_{j,t} + \xi_{j,t} \), where \( x_{j,t} \) is a vector of demand-related variables, \( k_{j,t} = \ln(K_{j,t}) \) represents the logarithm of the ways and structures capital stock of the railroad firm \( j \), \( p_{j,t} \) is the price of using the railroad firm \( j \) to provide the freight service, and \( \xi_{j,t} \) represents the unobservable efficiency of railroad firm \( j \) at time \( t \). The capital stock is updated at each time period using the relation \( K_{j,t} = K_{j,t-1}(1-d) + I_{j,t-1} \), where \( I_{j,t-1} \) represents the investment in the network infrastructure. I define the observed quality of firm \( j \) at time \( t \) as \( v_{j,t} = x_{j,t} \beta + \theta k_{j,t} \). This quality index will be used as a state variable in the dynamic model later on.

Using Berry (1994) and Train (1999), the formulas that characterize this nested logit model are:

\[
\begin{align*}
s_{j,g}(\delta, \sigma) &= \frac{\exp\left(\frac{\delta_{j,t}}{1-\sigma}\right)}{\sum_{j \in g} \exp\left(\frac{\delta_{j,t}}{1-\sigma}\right)} \text{ for railroad firm } j \text{ in group } g \text{ at time } t, \text{ and} \\
s_{g,t}(\delta, \sigma) &= \frac{D_{g,t}^{1-\sigma}}{\sum_{g} D_{g,t}^{1-\sigma}},
\end{align*}
\]

where \( s_{j,g}(\cdot) \) denotes the within market share of firm \( j \) at time \( t \) in the group \( g=1 \), \( \delta \) denotes the vector of mean-utilities of all railroad firms, \( \sigma \) denotes the within group correlation of railroad firms, and \( D_{g,t} = \sum_{j \in g} \exp(\delta_{j,t} / (1-\sigma)) \). Then, the market share of the outside alternative is given by \( s_{g,t}(\delta, \sigma) = 1/\sum_{g} D_{g,t}^{1-\sigma} \), and the market share for the railroad firm \( j \) can be expressed as:

\( \quad \)\(^8\) The mean-utility of the outside option is normalized to zero for each time period. This is a standard assumption in this type of demand model.
The consumer welfare can thus be computed as:

\[
CS_t = \frac{1}{\alpha} \ln \left( \sum_{g=0}^{1} D_{g,t}^{1-\sigma} \right) = \frac{1}{\alpha} \ln \left( 1 + D_{t}^{1-\sigma} \right).
\]

Following Berry (1994), for a particular railroad firm \( j \) at year \( t \), the estimating equations are:

\[
\ln s_{j,t} - \ln s_{0,t} = x_{j,t} \beta + \theta k_{j,t} - \alpha p_{j,t} + \sigma \ln s_{j,t,g} + \xi_{j,t}.
\]

When I do not take into account the selection bias, I specify the following conditional moment restriction:

\[
E \left[ \xi_{j,t} | z_{j,t} \right] = 0,
\]

for a set of instruments \( z_{j,t} \), and consistent estimates of the demand parameters are obtained through GMM. The instrumentation approach is standard in the literature. Instruments include the exogenous demand related variables that are included in the regression, the corresponding BLP (Berry, Levinsohn, and Pakes (1995)) instruments, and some cost-shifter variables.\(^9\)

However, as mentioned in the introduction, it is likely that there is selection on \( \xi \) due to firms with low values of \( \xi \) exiting. In this case, the moment condition (3) would in general be violated. This problem might be severe in the data since an important number of railroad firms exited the industry by selling their assets to other firms in the industry. Therefore, the condition (3) must be written as:\(^{10}\)

---

\(^9\) The variables used in the estimation are described in detail in the data appendix.

\(^{10}\) Using the moment conditions in (3) might lead to inconsistent estimates of the demand parameters, which will yield inconsistent estimates of the consumer welfare and inconsistent estimates of the marginal costs for firm \( j \) at time \( t \) (see Berry (1994) for the method to recover the marginal costs from the demand model).
where \( r_{j,t} \) is the selection indicator:

\[
 r_{j,t} = \begin{cases} 
 1 & \text{if firm } j \text{ is active at year } t, \\
 0 & \text{otherwise.} 
\end{cases}
\]

Formally, I write equation (2) as:

\[
\ln s_{j,t} - \ln s_{0,t} = x_{j,t} \beta + \theta k_{j,t} - \alpha p_{j,t} + \sigma \ln s_{j,t} + E[\xi_{j,t} | s_{j,t}, r_{j,t} = 1] + e_{j,t},
\]

where \( E[\xi_{j,t} | s_{j,t}, r_{j,t} = 1] = 0 \) by construction since \( e_{j,t} \equiv \xi_{j,t} - E[\xi_{j,t} | s_{j,t}, r_{j,t} = 1] \). To be able to estimate the equation (6), I need to compute the correction term for attrition \( E[\xi_{j,t} | s_{j,t}, r_{j,t} = 1] \). This requires to understand why the panel is unbalanced; in other words, why a firm is observed in the market \( (r_{j,t} = 1) \). The dynamic model of exit behavior, to which I now turn, will lead to compute the correction term since it will help understand why a firm decides to stay in (or exit from) the market.

### 3.2 The dynamic framework

I consider a model of dynamic competition between oligopolistic competitors. I build on Ericson and Pakes (1995), Berry, Ostrovsky, and Pakes (2007), and Dorazelski and Satterthwaite (2010). The key feature is that actions taken in a given period may affect both current profits and, by influencing a set of commonly observed state variables, future strategic interactions Thus, this model permits the aspects of dynamic competition such as exit and investment decisions. As mentioned in the introduction, I focus on the exit decision since it allows us to derive a correction term to deal with selection bias in the estimation of the demand model.

I consider infinitely lived agents that make decisions at times \( t = 1, \ldots, +\infty \), where \( J_t \) denotes the number of agents at time \( t \). Conditions at time \( t \) are summarized by a vector of state variables \( \mathbf{w}_t \in W = \mathbb{R}^{2J_t + 1} \), where \( \mathbf{w}_t = (J_t, \nu_{1,t}, \ldots, \nu_{J_t,t}, s_{1,t}, \ldots, s_{J_t,t}, \xi_{1,t}, \ldots, \xi_{J_t,t}) \), where \( \nu_{j,t} \) represents the
observable quality and $\xi_{j,t}$ represents the unobservable quality (or efficiency) of freight services provided by firm $j$ at the beginning of period $t$. Given the state $w_t$, firms choose their actions simultaneously. These actions include three components: a pricing decision, an investment decision, and an exit decision. In the model, the pricing decision is static, whereas investment and exit are dynamic decisions.

At the beginning of each period, each incumbent firm is assigned a random scrap value received upon exit. Scrap values are privately known, that is, whereas a firm learns its own scrap value prior to making its decisions, its rivals’ scrap values remain unknown to it. Adding firm heterogeneity in the form of randomly drawn and privately known scrap values leads to a game of incomplete information. An incumbent firm has to decide each period whether to remain in the industry and, if so, how much to invest. Once these decisions are made, product market competition takes place.

The model accounts for firm heterogeneity in two ways. First, all characteristics that are relevant to a firm’s profit from product market competition are encoded in its “state”. A firm is able to change its state over time through investment, although a higher investment today is not a guarantee for a more favorable state tomorrow, it does ensure a more favorable distribution over future states. In other words, the model allows that a firm’s transition from one state to another is subject to an idiosyncratic shock, thus there is variability in the fortunes of firms even if they carry out identical strategies. This heterogeneity, in the spirit of Ericson and Pakes (1995), determines the evolution of the state variables and will be interesting for the estimation of the investment policy function. However, this paper deals with selection bias due to exit and, as I show below, only the second type of heterogeneity in the scrap value is interesting for my purpose. Thus I do not enter into the details regarding the evolution of the state variables. Second, I introduce heterogeneity in firms by taking into account the differences in opportunity costs of staying in the industry. Indeed, I assume that incumbents have random scrap values that are independently and identically distributed across firms and over time.
The within period timing is the following:

- At date $t$, firms compete in the product market by making pricing decisions based on the vector of state variables, denoted $w_t$, where
  \[ w_t = (J, V_{j,t}, \ldots, V_{j,t}, \xi_{j,t}, \ldots, \xi_{j,t}) \];

- Each incumbent firm learns its scrap value, denoted $\phi_{j,t}$, and decides on exit and investment;

- At the end of period $t$, the vector of state variables is updated, and firms take their decisions in period $t+1$ conditional on $w_{t+1}$. Regarding the exit decisions, the number of active firms at period $t+1$ is equal to $J_t$ minus the number of firms that exit at period $t$. In addition, if the firm $j$ does not exit the market at period $t$, the investment decision is carried out and its state variables are updated.

I assume that at the beginning of the period each incumbent receives a random scrap value $\phi_{j,t}$ from a distribution $F(.)$ with $E(\phi_{j,t}) = \phi$, and the scrap values are identically and independently distributed across firms and over time. Incumbent firm $j$ at time $t$ learns its scrap value prior to its exit and investment decisions, but the scrap value of its rivals remain unknown to it. Let $\chi_{j,t}(w_t, \phi_{j,t}) = 1$ indicates that the incumbent firms chooses to stay in the industry in state $w_t$, and $\chi_{j,t}(w_t, \phi_{j,t}) = 0$ denotes the firm $j$ chooses to exit the industry at time $t$. Because this decision is conditioned on its private scrap value, it is a random variable from the perspective of the other firms. I use $\xi_{j,t}(w_t) = \int \chi_{j,t}(w_t, \phi_{j,t}) dF(\phi_{j,t})$ to denote the probability that firm $j$ at date $t$ remains active in the industry in state $w_t$.

The firms compete in prices in the product market. The two state variables of a firm $j$ at time $t$ are denoted $w_{j,t} = (V_{j,t}, \xi_{j,t})$. The optimal price is a function of the quality level of the firm $j$, $w_{j,t}$, and the quality levels of its competitors $w_{-j,t} = (w_{i,j-1,t}, w_{j-1,t}, w_{j+1,t}, \ldots, w_{j,t})$. Thus, the profit
function from the product market competition for firm \( j \) at time \( t \) is a function of the industry state \( w_t = (J_t; w_{j,t}; w_{-j,t}) \), \( \pi_{j,t}(w_t) \).

In addition to receiving profit, the active firm incurs the cost of investment, \( c(i_{j,t}) \), where the investment level \( i_{j,t} \) is decided at the beginning of the period. The active firm moves then to the next period state according to the transitions of the state variables.

Let us denote by \( V_{j,t}(w_t,\phi_{j,t}) \) the expected net present value of all future profits for the firm \( j \) at date \( t \), which is defined recursively by the solution of the Bellman equation:

\[
V_{j,t}(w_t,\phi_{j,t}) = \max_{\pi_{j,t}} \left\{ \pi_{j,t}(w_t) + (1 - \chi_{j,t}(w_t,\phi_{j,t}))\phi_{j,t} + \chi_{j,t}(w_t,\phi_{j,t}) \right\}
\]

\[
\times \left\{ -c(i_{j,t}(w_t)) + \delta E \left[ V_{j,t+1}(w_{t+1},i_{j,t+1}(w_t),i_{-j,t+1}(w_t),\phi_{-j,t+1}(w_t)) \right] \right\},
\]

where \( \delta \) represents the discount rate. The expression \( V_{j,t}(w_t,\phi_{j,t}) \) denotes the value function after the firm has drawn its scrap value, and \( V_{j,t}(w_t) = \int V_{j,t}(w_t,\phi_{j,t}) \, dF(\phi_{j,t}) \) denotes the value function before the firm has drawn its scrap value. If the firm stays in the industry, it gets the profit from the product-market competition plus the continuation value minus the cost of investment. If the firm leaves the industry, it gets the profit from the product-market competition plus the scrap value of exit. By comparing the value of staying (i.e. the continuation value minus the cost of investment) with the scrap value of exiting, a firm takes the decision to exit (i.e. to sell its assets to another firm in the industry, and gets a scrap value from a resale market). Note that it takes one time-period to implement the exit decision. A firm that takes the exit decision at period \( t \) gets the profit from the product market and effectively leaves the market at the end of period \( t \).

From (7), the optimal decision of firm \( j \) at date \( t \) to remain in the industry is a cut-off rule characterized by:
\[
\chi_{j,t}(w_t, \phi_{j,t}) = \begin{cases} 
1 & \text{if } \phi_{j,t} < \bar{\phi}_{j,t}(w_t), \\
0 & \text{if } \phi_{j,t} > \bar{\phi}_{j,t}(w_t), 
\end{cases}
\] (8)

where \( \bar{\phi}_{j,t}(w_t) = \sup_{i_t} c(i_{j,t}(w_t)) + \delta E \{ V_{t+1} | w_t, i_{j,t}(w_t), i_{-j,t}(w_t), \xi_{-j,t}(w_t) \} \). Assuming that \( \phi_{j,t} \) follows a normal distribution \( N(\phi, \omega) \), I can write (8) as:

\[
\chi_{j,t}(w_t, \phi_{j,t}) = \begin{cases} 
1 & \text{if } \frac{\phi_{j,t} - \phi}{\omega} < \frac{\bar{\phi}_{j,t}(w_t) - \phi}{\omega}, \\
0 & \text{otherwise}, 
\end{cases}
\] (9)

Thus, the decision rule \( \chi_{j,t}(w_t, \phi_{j,t}) = 1 \) can be written equivalently \( 1(\phi_{j,t} < \bar{\phi}_{j,t}(w_t)) \) or

\[
1 \left( \frac{\phi_{j,t} - \phi}{\omega} < \frac{\bar{\phi}_{j,t}(w_t) - \phi}{\omega} \right),
\]

where \( 1(.) \) represents the indicator function. This point is crucial for the estimation algorithm to correct for selection bias in the demand equation. Then the probability of staying in the market can be written as:

\[
\zeta_{j,t}(w_t) = \int \chi_{j,t}(w_t, \phi_{j,t}) dF(\phi_{j,t}) = \int \left[ \frac{\phi_{j,t} - \phi}{\omega} < \frac{\bar{\phi}_{j,t}(w_t) - \phi}{\omega} \right] dF(\phi_{j,t}) = F \left( \frac{\bar{\phi}_{j,t}(w_t) - \phi}{\omega} \right),
\] (10)

where \( F(.) \) is the cumulative distribution function of the standard normal distribution.

Next the threshold value such that a firm stays or exits the market is obtained using the expression:

\[
F^{-1} \left( \zeta_{j,t}(w_t) \right) = \frac{\bar{\phi}_{j,t}(w_t) - \phi}{\omega}.
\] (11)

Assuming the state of the industry \( w_t \) is observed, the probability of remaining in the industry, \( \hat{\zeta}_{j,t}(w_t) \), can be estimated for all \((j,t)\) (see the next section for details on the estimation algorithm). Next, inverting the standard normal distribution and evaluating at \( \hat{\zeta}_{j,t}(w_t) \), the threshold value that
determines if a firm decides to stay in the market can be computed, \( (\bar{\phi}_{j,t}(w_t) - \phi) / \omega \). This threshold will allow computing a correction term in order to take into account the selection bias in the estimation of the demand model.

Before detailing the estimation algorithm, it is important to emphasize the link between the demand and the dynamic exit model. I have defined the observed quality of firm \( j \) at time \( t \) as \( v_{j,t} = x_{j,t} \beta + \theta k_{j,t} \), where \( v_{j,t} \) denotes the observed state variable for the firm \( j \) at date \( t \) in the dynamic model. The second state variable \( \xi_{j,t} \), which is unobservable by the econometrician, is directly incorporated into the demand model. The definitions of the state variables \( w_{j,t} = (v_{j,t}, \xi_{j,t}) \) imply that the spot profit function can be written as \( \pi_{j,t}(w_t) \), where \( w_t = (J_t, w_{1,t}, \ldots, w_{J,t}, \ldots, w_{J,t}) \).

This implies that the value function also depends on the same state variables, \( V_{j,t}(w_t) \). This connection between the dynamic model and the demand model is essential to have a coherent framework.

### 4 Demand model and selection: econometric methodology

I present the estimation algorithm that allows us to control for the selection effect in the demand equation (6) using the dynamic model of firm behavior. I also explicitly discuss the issues linked to the exogenous process of the unobserved state variable when a panel data is characterized by a self selection effect. In some sense, this discussion can be related to Wooldridge (1995) and Semykina and Wooldridge (2005) since I include firm fixed-effects in the analysis.

From equation (6), the correction term for the selection bias is \( E[\xi_{j,t} | x_{j,t}, r_{j,t} = 1] \). From the dynamic model, it takes one time-period to implement the exit decision. Thus the firm is active at time \( t \) if it decides to stay on the market in the previous period \( t - 1 \), that is if \( \phi_{j,t-1} < \bar{\phi}_{j,t-1}(w_{t-1}) \). Thus, to
be able to compute the correction term I need to condition with respect to the previous state of the industry \( w_{t-1} \). This leads to the following correction term:

\[
E[\xi_{j,t} \mid z_{j,t}, w_{t-1}, r_{j,t} = 1] = E[\xi_{j,t} \mid z_{j,t}, w_{t-1}, \chi_{j,t-1} = 1] = E[\xi_{j,t} \mid z_{j,t}, w_{t-1}, \phi_{j,t-1} < \phi_{j,t-1}(w_{t-1})].
\] (12)

It means that, in the presence of attrition, the object of interest is \( E[y_{j,t} \mid z_{j,t}, w_{t-1}, r_{j,t} = 1] \), where \( y_{j,t} \) denotes the dependent variable \( \ln s_{j,t} - \ln s_{0,t} \).

Thus, the estimating equation can be written as follows:

\[
y_{j,t} = x_{j,t} \beta + \theta k_{j,t} - \alpha p_{j,t} + \sigma \ln s_{j,t} + E[\xi_{j,t} \mid z_{j,t}, w_{t-1}, r_{j,t} = 1] + e_{j,t},
\] (13)

where \( e_{j,t} = \xi_{j,t} - E[\xi_{j,t} \mid z_{j,t}, w_{t-1}, r_{j,t} = 1] \). The moment conditions are derived using \( E[e_{j,t} \mid z_{j,t}, w_{t-1}, r_{j,t} = 1] = 0 \).

For the following, the correction term for attrition is computed using the law of iterated expectation:

\[
E[\xi_{j,t} \mid z_{j,t}, w_{t-1}, r_{j,t} = 1] = E[E(\xi_{j,t} \mid z_{j,t}, w_{t-1}, \phi_{j,t-1}) \mid z_{j,t}, w_{t-1}, r_{j,t} = 1].
\] (14)

The exact form of the correction term depends on the specification of the process for the unobserved firm efficiency \( \xi_{j,t} \). The following section presents the estimation algorithm depending on the process of the unobserved firm efficiency \( \xi_{j,t} \).

4.1 Estimation algorithm: firm fixed-effect and attrition

I assume the following specification for the unobserved efficiency:

\[
\xi_{j,t} = e_{j} + \rho \phi_{j,t-1} + \tau_{j,t},
\] (15)
where \( c_j \) denotes the firm fixed-effect, \( \phi_{j,t-1} \) denotes the scrap value, and \( \tau_{j,t} \) denotes an error term independent from \( c_j \) and \( \phi_{j,t-1} \). This process means that private information last period, \( \phi_{j,t-1} \), becomes public information the next period, \( \xi_{j,t} \).

The inclusion of a firm-fixed effect can be related to the work of Wooldridge (1995) and Semykina and Wooldridge (2005). Indeed, incorporating firm fixed-effect is attractive when we suspect firms to select out of the sample based on unobserved fixed heterogeneity. This can also be related to Nevo (2000, 2001) which advocates the use of fixed-effect to ensure that the observed characteristics capture the true factors that determine utility and it improves the fit of the model.

The correction term can be written as:

\[
E[\xi_{j,t} \mid z_j, w_{t-1}, r_j] = E(E(\xi_{j,t} \mid z_j, w_{t-1}, \phi_{j,t-1}) \mid z_j, w_{t-1}, r_j] \\
= c_j + \rho E[\phi_{j,t-1} \mid w_{t-1}, r_j] \\
= c_j + \rho E[\phi_{j,t-1} \mid w_{t-1}, r_{j,t} = 1] \\
= c_j + \rho E[\phi_{j,t-1} \mid w_{t-1}, \phi_{j,t-1} < \bar{\phi}_{j,t-1}(w_{t-1})] \\
= c_j + \rho \lambda_j(w_{t-1}),
\]

where \( \lambda_j(w_{t-1}) \) denotes the Mills ratio which is correcting for attrition, that is:

\[
\lambda_j(w_{t-1}) = \frac{f\left(\bar{\phi}_{j,t-1}(w_{t-1})\right)}{F\left(\bar{\phi}_{j,t-1}(w_{t-1})\right)},
\]

and \( f(.) \) and \( F(.) \) represents respectively the probability distribution function and the cumulative distribution function of the standard normal distribution (indeed, in the dynamic model, I have assumed that the scrap value \( \phi_{j,t-1} \) follows a normal distribution \( N(0, \omega) \)). See appendix for a robustness analysis with another distribution for the scrap value. Moreover, in equation (16) it is observed that I condition with respect to \((z_j, r_j)\) and no more with respect to \((z_{j,t}, r_{j,t} = 1)\). This is due to the inclusion of the fixed-effects in the analysis (see Wooldridge (1995, 2002, 2005)).
assumed that $E[\phi_{j,t-1} | w_{t-1}, r_j] = E[\phi_{j,t-1} | w_{t-1}, r_{j,t}]$. This is coherent with the dynamic exit model since the scrap value is independent over time.\footnote{Without this assumption, the term $E[\phi_{j,t-1} | w_{t-1}, r_j]$ would be very complicated to compute (see Wooldridge (1995, 2002), and Semykina and Wooldridge (2005)).} Using (16), the dynamic exit model provides a guide for the construction of the attrition correction term in a structural econometrics framework. It is like a Heckman two-step procedure where the correction term comes from the economic theory using a dynamic model of exit behavior.

Then the following equation is obtained:

$$\ln s_{j,t} - \ln s_{o,j} = x_{j,t} \beta + \theta k_{j,t} - \alpha p_{j,t} + \sigma \ln s_{j,t|g=1} + \epsilon_j + \rho \lambda_{j,t-1}(w_{t-1}) + e_{j,t}, \quad (18)$$

where $E[e_{j,t} | z_j, w_{t-1}, r_j] = 0$ by construction. Note that the variance of the scrap value is normalized to one, $\omega = 1$. Indeed, in equation (18) I can identify only the parameters $\rho \omega$ and not each parameter separately. For the following, I denote $\lambda_{j,t-1}(w_{t-1})$ by $\lambda_{j,t-1}$, and $\ln s_{j,t} - \ln s_{o,j}$ by $y_{j,t}$. I eliminate the firm fixed-effect by a first-difference, which leads to the following estimating equation:

$$\Delta y_{j,t} = \Delta x_{j,t} \beta + \theta \Delta k_{j,t} - \alpha \Delta p_{j,t} + \sigma \Delta \ln s_{j,t|g=1} + \rho \Delta \lambda_{j,t-1} + \Delta e_{j,t}, \quad (19)$$

where $\Delta$ denotes the first-difference operator: $\Delta e_{j,t} = e_{j,t} - e_{j,t-1}$. It is important to emphasize the choice of instruments for the equation (19). The obvious instruments are the exogenous variables $z_{j,t}$ and $z_{j,t-1}$ (the exogenous demand variables $x$ are included in the instruments $z$). I follow Vellturo (1989) and Berndt et al. (1993a, 1993b) in using exogenous demand-related variables, denoted $x_{j,t}$, which can be constructed on a firm-specific basis. The variable —coal consumption ($CCON$)— is measured on a state-by-state basis and then aggregated across states to be railroad specific and to conform to each railroads’ operating territory. These aggregations vary from year to year as some railroad firms exit the industry and some other railroad firms extend their networks by buying the assets of the firms that exit the industry. The use of the variable $CCON$ seems justified since coal is the
main commodity carried by the US railroad firms (see section 2).\textsuperscript{13} I also include a quadratic time trend in order to capture any unexplained productivity growth. The instruments include the strictly exogenous variables: the coal consumption (CON), the quadratic time trend, the BLP instrument of the coal consumption, the miles of road operated (ROAD), the average length of haul (HAUL), the lag of these six variables, and the lag of the BLP instrument for ROAD. The two variables ROAD and HAUL are used as instruments since they are considered as cost-shifters (see Berndt et al (1993a, 1993b), Ivaldi and McCullough (2001, 2008)). The lags of these variables can also be used as instruments since the equation (19) is estimated in first-difference. More details are given in the data appendix regarding the construction of the instruments.

Let us now discuss the endogeneity of the other variables included in the estimating equation (19). The price, denoted $p_{j,t}$, and the within market share, denoted $\ln s_{j,t|g}$, are endogenous. Thus, the variables $\Delta p_{j,t}$ and $\Delta \ln s_{j,t|g=1}$ are also endogenous. The discussion becomes a little more subtle for the variables $\Delta k_{j,t}$ and $\Delta \lambda_{j,t-1}$. Using the structure of the model, I know that the variables $k_{j,t}$ and $\lambda_{j,t-1}$ are weakly exogenous. Indeed, $k_{j,t} = \ln(K_{j,t})$, and the capital stock is constructed using the relation $K_{j,t} = K_{j,t-1}(1-\delta) + I_{j,t-1}$, where $I_{j,t-1}$ represents the investment in the network at date $t-1$ (more details regarding the construction of the capital stock are available in the data appendix). From the dynamic model, I know that the investment, $I_{j,t-1}$, is endogenous and it is a function of the previous state of the industry, $w_{t-1}$. This implies that the capital stock $K_{j,t}$, and thus the proxy for network quality $k_{j,t}$, are a function of $w_{t-1}$. By construction, the error term $e_{j,t}$ in equation (18) is uncorrelated with the previous state of the industry $w_{t-1}$. Thus, the proxy for the network quality, $k_{j,t}$, is weakly exogenous since it is uncorrelated with the contemporaneous and the future error terms, $e_{j,s}, s \geq t$, and correlated with the past error term, $e_{j,t}, s \leq t-1$. This implies that in the estimating equation (19), the variable $\Delta k_{j,t} = k_{j,t} - k_{j,t-1}$ is endogenous since $k_{j,t}$ is correlated with $\Delta e_{j,t}$.

\textsuperscript{13} If a railroad firm is crossing some states with high coal consumption, then it is likely that it will face a high demand for its freight services since coal is the main commodity carried by the US railroad firms.
through $e_{j,t-1}$. Nevertheless, I can instrument $\Delta k_{j,t}$ by using the $K_{j,t-1}$ as instrument since the lag of the capital stock is a function of the state of the industry at date $t-2$, $w_{t-2}$, and the error term is $\Delta e_{j,t}$ is uncorrelated with the state of the industry at date $t-2$ (for the estimation, I have also added $K_{j,t-2}$ as an instrument). Lastly, I discuss the endogeneity of the first-difference of the Mills ratio, $\Delta \lambda_{j,t-1} = \lambda_{j,t-1}(w_{t-1}) - \lambda_{j,t-2}(w_{t-2})$. Like the stock of capital, the Mills ratio $\lambda_{j,t-1}(w_{t-1})$ is also weakly exogenous since it is uncorrelated with $e_{j,s}, s \geq t$ and it is correlated with $e_{j,s}, s \leq t-1$. In the estimating equation (19), $\Delta \lambda_{j,t-1}$ is endogenous since $\lambda_{j,t-1}$ is correlated with $e_{j,t-1}$ and thus with $\Delta e_{j,t}$. I instrument $\Delta \lambda_{t-1} = \lambda_{j,t-1} - \lambda_{j,t-2}$ by the second lag of the Mills ratio, that is $\lambda_{j,t-2}$.

To summarize, the choice of the instruments is guided by the structure of the model. Hence, during the estimation, accepting the over-identifying restriction may be interpreted as accepting the structure of the model as well. I will come back on this issue when I present the estimation results in section 5.

In the estimating equation (19), I have assumed that the previous state of the industry is known, $w_{t-2}$, since I use the conditioning $E[\Delta e_{j,t} \mid z_j, w_{t-2}, r_j] = 0$. To make the estimation feasible, I need to use the following iterative algorithm:

1) Start with an initial guess of the vector of demand parameters, denoted $\hat{\mu} = (\hat{\beta}, \hat{\alpha}, \hat{\sigma})$.

2) Using the equation (2), I compute an estimate of the unobserved state variable that represents the unobserved firm efficiency, $\hat{\xi}_{j,t}$, and I compute the observed state variable of each firm, $\hat{\nu}_{j,t} = x_{j,t} \hat{\beta} + \hat{\theta} k_{j,t}$, $\forall (j,t)$.

3) I compute the probabilities of remaining in the industry as a function of the industry state, $\hat{\xi}_{j,t}(\hat{w}_t)$, where $\hat{w}_t = (J_t; \hat{\nu}_{j,t}, \hat{\nu}_{-j,t}, \hat{\xi}_{j,t}, \hat{\xi}_{-j,t})$, using a probit model, and $\hat{\nu}_{-j,t}$ and $\hat{\xi}_{-j,t}$ represent respectively the sum of the observed and the unobserved state variable for the competitors. The threshold value $\frac{\tilde{\phi}_{j,t-1}(\hat{w}_{t-1})}{\omega} = F^{-1}(\tilde{\xi}_{j,t-1}(\hat{w}_{t-1}))$ is computed. This can be related to Hotz and Miller (1993). They show the existence of an invertible mapping.
between conditional choice probabilities and the continuation values. This enables to obtain the continuation values nonparametrically first by estimating the agent’s choice probabilities at a given state and then inverting the choice probabilities to obtain the relevant continuation values. In my empirical application, a nonparametric estimate of the conditional choice probabilities can be very imprecise due to the small sample size. Thus, I use a second-order polynomial approximation in \( (J_t; \hat{V}_{j,t}, \hat{V}_{-j,t}; \hat{\xi}_{j,t}, \hat{\xi}_{-j,t}) \) to evaluate the conditional choice probabilities.

4) Using the recovered threshold \( \frac{\hat{\phi}_{j,t-1}(\hat{w}_{t-1})}{\omega} \), I obtain the Mills ratio \( \hat{\lambda}_{j,t-1}(\hat{w}_{t-1}) \) as a correction term for attrition (see equation (17)). I am also able to recover \( \hat{\lambda}_{j,t-2}(\hat{w}_{t-2}) \).

5) I estimate the regression (19) by GMM using the instruments \( z_{j,t}, z_{j,t-1}, K_{j,t-1}, K_{j,t-2}, \) and \( \hat{\lambda}_{j,t-2} \).

6) Using the new demand estimates \( \hat{\mu} = (\hat{\beta}, \hat{\phi}, \hat{\alpha}, \hat{\sigma}) \), I repeat steps 2-5 until convergence of the demand estimates.

The key point in the estimation algorithm is the assumption on the exogenous process of the unobserved firm efficiency \( \xi_{j,t} \) which is necessary to compute the correction term in equations (18) and (19). However, the presence of serial correlation can be a concern. In particular, this has an impact on the correction term \( E[\xi_{j,t} | z_j, w_{t-1}, r_j] = E[E(\xi_{j,t} | z_j, w_{t-1}, \phi_{j,t-1}) | z_j, w_{t-1}, r_j] \), since \( \xi_{j,t-1} \) is part of the previous state of the industry \( w_{t-1} \). To deal with this issue, the next subsection explicitly includes serial correlation in the unobserved firm efficiency.

### 4.2 Estimation algorithm: firm fixed-effect, serial correlation, and attrition

I assume the following exogenous process for the unobserved efficiency:

\[
\xi_{j,t} = c_j + \gamma \xi_{j,t-1} + \rho \phi_{j,t-1} + \tau_{j,t},
\] (20)

where the error term \( \tau_{j,t} \) is identically and independently distributed. The correction term can be written as:
\[ E(\xi_{jt} \mid z_j, w_{t-1}, r_j) = E(E(\xi_{jt} \mid z_j, w_{t-1}, \phi_j, \phi_{t-1}) \mid z_j, w_{t-1}, r_j) = c_j + \gamma \xi_{j,t-1} + \rho \lambda (w_{t-1}). \] (21)

The equation (18) becomes:
\[ \ln s_{j,t} - \ln s_{o,t} = x_{j,t} \beta + \theta k_{j,t} - \alpha \ln s_{j,t,g=1} + c_j + \gamma \xi_{j,t-1} + \rho \lambda_{j,t-1} + e_{j,t}, \] (22)

I eliminate the fixed-effect by a first-difference which leads to the following estimating equation:
\[ \Delta y_{j,t} = \Delta x_{j,t} \beta + \theta \Delta k_{j,t} - \alpha \Delta \ln s_{j,t,g=1} + \gamma \Delta \xi_{j,t-1} + \rho \Delta \lambda_{j,t-1} + \Delta e_{j,t}, \] (23)

\[
\text{where } E[\Delta e_{j,t} \mid z_j, r_j, w_{t-2}] = 0 \text{ by construction. To estimate this equation, I replace } \Delta \xi_{j,t-1} \text{ by } \\
\Delta y_{j,t-1} - \Delta x_{j,t-1} \beta - \theta \Delta k_{j,t-1} + \alpha \Delta \ln s_{j,t-1,g=1} - \sigma \Delta \ln s_{j,t-1,g=1} \text{ (see equation (2)). The coefficient in front of } \\
\Delta y_{j,t-1}, \text{ denoted } \gamma, \text{ represents the importance of serial correlation. The instruments used here are the same as before.}
\]

4.3 Estimation algorithm: no attrition

The purpose of this section is to show that if attrition does not matter, then the iterative algorithm vanishes to a simple instrumental variable estimation.

Assume the following process for the unobserved firm efficiency:
\[ \xi_{j,t} = c_j + \gamma \xi_{j,t-1} + \tau_{j,t}. \] (24)

Then, the correction term is equal to:
\[ E[\xi_{j,t} \mid z_j, w_{t-1}, r_j] = c_j + \gamma \xi_{j,t-1}, \] (25)

which implies the following estimating equation:
\[ y_{j,t} = x_{j,t} \beta + \theta k_{j,t} - \alpha \ln s_{j,t,g=1} + c_j + \gamma \xi_{j,t-1} + e_{j,t}, \] (26)
where \( E[e_{j,t} \mid z_{j,t}, w_{t-1}, r_j] = 0 \) by construction and \( y_{j,t} = \ln s_{j,t} - \ln s_{0_j} \). In other words, the object of interest can be defined as \( E[y_{j,t} \mid z_{j,t}, w_{t-1}, r_j] = 0 \). Since I do not include a Mills ratio to correct for attrition, it is not necessary to condition with respect to the previous state of the industry, \( w_{t-1} \). Thus, I consider the following object:

\[
E[y_{j,t} \mid z_{j,t}, r_j] = E[ E[y_{j,t} \mid z_{j,t}, r_j, w_{t-1}] \mid z_{j,t}, r_j].
\] (27)

Replacing \( y_{j,t} \) by its expression in (26), I obtain:

\[
E[y_{j,t} \mid z_{j,t}, r_j] = E[ x_{j,t} \beta + \theta k_{j,t} - \alpha p_{j,t} + \sigma \ln s_{j,t+1} + c_j + \gamma E[\xi_{j,t-1} \mid z_{j,t}, r_j] ]
\]

\[
= x_{j,t} \beta + \theta k_{j,t} - \alpha p_{j,t} + \sigma \ln s_{j,t+1} + c_j + \gamma E[\xi_{j,t-1} \mid z_{j,t}, r_j]
\]

\[
= x_{j,t} \beta + \theta k_{j,t} - \alpha p_{j,t} + \sigma \ln s_{j,t+1} + c_j,
\] (28)

where I assume that \( E[\xi_{j,t-1} \mid z_{j,t}, r_j] = 0 \). The estimation of equation (28) is standard and the iterative algorithm vanishes to a standard regression. I estimate it by first-differencing to remove the firm fixed-effect and I use the same instruments as in the previous subsections. More precisely, I use the exogenous variables \( z_{j,t} \) and \( z_{j,t-1} \). I also use the previous capital stock \( K_{j,t-1} \) and \( K_{j,t-2} \) as instruments.

The last case that I consider is:

\[
\xi_{j,t} = c_j + \tau_{j,t}.
\] (29)

This implies that I do not consider attrition and serial correlation in the exogenous process of the unobserved firm efficiency. With respect to the previous case, I add the current capital stock, \( K_{j,t} \), as an instrument. Indeed, if I assume that there is not serial correlation and no attrition in the process of the unobserved firm efficiency, the structure of the model tells us to include this variable as an instrument.
5 Estimation results

This section presents the estimation results of the demand model using the algorithm to correct for selection due to attrition of firms. The crucial point in the estimation method to correct for selection bias is to derive a cut-off value, denoted $\phi_{j,t}$, that determines the exit/staying decision.\textsuperscript{14} The equations (10) and (11) are the bases of the estimation strategy. Indeed, in the estimation algorithm, I have to compute only the probability of staying in the industry in order to correct for the selection bias due to exit. The advantage of the methodology is to correct for the attrition bias using a dynamic model to derive a selection equation, but I do not need to solve the dynamic model to obtain the correction term. The method is relatively simple and the only computational burden might concern the iterative process but convergence was very quick to achieve during the estimations.

Table 1 reports the estimation results using four different methodologies:

- In column 1, I implement a standard fixed-effect analysis with no serial correlation and no attrition (see equation (29));
- In column 2, I implement a fixed-effect analysis by correcting for the selection bias (see equation (15));
- In column 3, I implement a fixed-effect analysis by taking into account serial correlation but not attrition (see equation (24));
- In column 4, I implement a fixed-effect analysis by correcting for attrition and by taking into account the potential serial correlation in the unobserved firm efficiency (see equation (20)).

The results from column 1 in Table 1 is puzzling since it implies that an increase in the quality of a network has a negative impact on the utility of the consumers (the coefficient of the capital stock is negative, $-0.0645$, which is counter-intuitive). This negative estimate may come from a selection bias as Olley and Pakes (1996) mentioned in their paper on production function. As column 2 shows, correcting for selection implies a positive impact of the capital stock on demand (+0.3010). The

\textsuperscript{14}Olley and Pakes (1996) use a similar idea in the context of production function, with a constant scrap value over time and no strategic interactions between firms. Their threshold is determined as a cut-off value in term of productivity.
intuition of this result follows from (16). I know that \( E(\phi_{j,t-1} | \phi_{j,t-1} < \phi_{j,t-1}(w_{t-1})) \) is equal to
\[
- f(\phi_{j,t-1})/ F(\phi_{j,t-1}),
\]
where I normalized the variance \( \omega \) of the scrap value to one. I can check that this expectation is negative for \( \phi_{j,t} \) positive and this is the case in the model. Indeed, I recover the threshold \( \phi_{j,t} \) by evaluating the inverse of the normal cumulative distribution function in the probability of staying in the industry; since the estimated probabilities, denoted \( \hat{\phi}_{j,t} \), are always above 0.5, it implies that the threshold are positive, \( \phi_{j,t} > 0 \forall (j,t) \). Moreover, I know that
\[
E(\xi_{j,t} | z_{j,t}, w_{t-1}, r_{j}) = c_{j} + \rho E(\phi_{j,t-1} | z_{j,t}, w_{t-1}, \phi_{j,t-1} < \phi_{j,t-1}) = c_{j} + \rho \lambda_{j,t-1},
\]
where the parameter \( \rho \) represents the correlation between the unobserved state variable \( \xi_{j,t} \) and the value of exit \( \phi_{j,t} \). I expect \( \rho > 0 \) since it is reasonable to think that a higher value of exit is associated with a higher value of the unobserved state variable. This implies that the correction term for attrition is negatively correlated with the stock of capital and I have a negative bias when I do not take into account the selection effect in column 1. When I control for attrition in column 2, the estimation algorithm allows removing this negative bias and I find a positive impact of the quality of the network on the utility of the consumers (+0.3010). I can also illustrate the selection bias by comparing column 3 and column 4. In the presence of serial correlation, when I do not take into account the attrition, I underestimate by 70\% the coefficient of the network quality (0.07 in column 3 instead of 0.25 in column 4).

Comparing column 1 and column 2, I also see that the price coefficient is under estimated when I do not correct for attrition (38.91 in column 1 instead of 44.63 in column 2). Moreover, the comparison of column 3 and column 4 also illustrates the selection bias. Specifically, when I control

---

15 Intuitively, increasing the proxy for network quality leads to an increase in the probability of staying. This implies an increase in the threshold value of staying, \( \phi_{j,t} \), and thus a decrease in the Mills ratio \( \lambda_{j,t} \). Since \( \rho \) is positive, it implies that the correction term for attrition, \( \rho \lambda_{j,t} \), is also decreasing. Thus, the correction term for attrition is negatively correlated with the proxy for network quality and this causes a negative bias in column 1 regarding the coefficient of the network quality. The same reasoning applies for the coefficients of the variables \( x_{j} \) since they are included in the observed state variable \( \nu_{j,t} = x_{j,t} \beta + \theta k_{j,t} \). However, in Table 1, I see that the coefficients of the coal consumption and of the quadratic time trend are roughly constant across the four specifications.
for serial correlation, but not correct for the attrition, the price coefficient is still underestimated (39.56 in column 3 instead of 66.26 in column 4).

To summarize, I see that attrition may create some important biases in the coefficient estimates (which might lead to counter-intuitive results). The estimation algorithm controls for the attrition/selection bias and column 2 shows that attrition is significant at a 5% level. However, as mentioned in the subsection 4.2, some bias can remain in the estimates of column 2 due to serial correlation in the unobserved firm efficiency. For example, I can prove that the coefficient of the correction term $\rho$ is under-estimated when I do not take into account the potential serial correlation (the reasoning is similar to the previous one regarding the negative bias for the coefficient of the network quality). This is coherent with column 4 in Table 1 which shows an increase in the coefficient of the correction term for attrition (0.6795). Furthermore, the coefficient of the within market share, denoted $\sigma$, is decreasing when I take into account attrition (0.9582 in column 1 instead of 0.4899 in column 2; 0.6994 in column 3 instead of 0.6035 in column 4).\footnote{The coefficient of the within market share, $\sigma$, is almost significant at 10% in column 3 and column 4.}

Lastly, the Sargan-test is the best for the column 4 when I control for attrition and serial correlation. This means that I accept the over-identifying restrictions for column 4, and in some sense, it also means that I accept the structure of the model (see section 4). By comparing only the columns 1, 2, and 3, I also see that the best Sargan test corresponds to column 2 where I control for attrition.

Next, I report the evolutions of the consumer surplus and the evolution of the marginal costs of firms using the parameter estimates of column 4. I choose this particular model since it controls for attrition and serial correlation in the unobserved firm efficiency. Moreover, the Sargan test is the best and does not reject the over-identifying restrictions (see Table 1).
Table 1. Demand estimates

<table>
<thead>
<tr>
<th></th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
<th>Case 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>attrition</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>serial correlation</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>price ($-\alpha$)</td>
<td>-38.9059**</td>
<td>-44.6331***</td>
<td>-39.5622**</td>
<td>-66.2602***</td>
</tr>
<tr>
<td></td>
<td>(15.4278)</td>
<td>(15.7474)</td>
<td>(15.2709)</td>
<td>(20.3032)</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.8582***</td>
<td>0.4899</td>
<td>0.6994**</td>
<td>0.6035</td>
</tr>
<tr>
<td></td>
<td>(0.2339)</td>
<td>(0.3181)</td>
<td>(0.3409)</td>
<td>(0.3980)</td>
</tr>
<tr>
<td>serial corr term: $\gamma$</td>
<td>NA</td>
<td>NA</td>
<td>Not estimated</td>
<td>0.1686</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(see eq. (28))</td>
<td>(0.1948)</td>
</tr>
<tr>
<td>correction term: $\rho$</td>
<td>NA</td>
<td>0.4611**</td>
<td>NA</td>
<td>0.6795**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.1990)</td>
<td></td>
<td>(0.3181)</td>
</tr>
<tr>
<td>$k_{jt}$</td>
<td>-0.0645</td>
<td>0.3010</td>
<td>0.0746</td>
<td>0.2569</td>
</tr>
<tr>
<td></td>
<td>(0.0889)</td>
<td>(0.2209)</td>
<td>(0.2291)</td>
<td>(0.2371)</td>
</tr>
<tr>
<td>CCON</td>
<td>0.0009</td>
<td>0.0006</td>
<td>0.0008</td>
<td>0.0005</td>
</tr>
<tr>
<td></td>
<td>(0.0009)</td>
<td>(0.0010)</td>
<td>(0.0009)</td>
<td>(0.0016)</td>
</tr>
<tr>
<td>time</td>
<td>-0.0858***</td>
<td>-0.0537</td>
<td>-0.0772**</td>
<td>-0.0957**</td>
</tr>
<tr>
<td></td>
<td>(0.0313)</td>
<td>(0.0337)</td>
<td>(0.0364)</td>
<td>(0.0501)</td>
</tr>
<tr>
<td>time squared</td>
<td>0.0023***</td>
<td>0.0015**</td>
<td>0.0022***</td>
<td>0.0023***</td>
</tr>
<tr>
<td></td>
<td>(0.0006)</td>
<td>(0.0006)</td>
<td>(0.0007)</td>
<td>(0.0010)</td>
</tr>
<tr>
<td>N</td>
<td>303</td>
<td>298</td>
<td>303</td>
<td>298</td>
</tr>
<tr>
<td>deg freedom</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Sargan</td>
<td>31.4801</td>
<td>14.0787</td>
<td>28.8672</td>
<td>5.5063</td>
</tr>
<tr>
<td>p value</td>
<td>0.0002</td>
<td>0.0797</td>
<td>0.0003</td>
<td>0.5984</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01,

As it is standard in the literature, I recover the marginal costs for each firm at a particular year by assuming that the firms compete in Bertrand. Using the first-order condition

$$(p_{jt} - mc_{jt}) \frac{\partial s_{jt}}{\partial p_{jt}} + s_{jt} = 0,$$

17 Note: The number of observations decreases from 353 to 298/303 due to the lag in the correction term and due to the use of lag instruments. Standard errors (in parenthesis) are computed by cluster-bootstrap.
I obtain the following formula for the mark-up:

\[
p_{jt} - mc_{jt} = \frac{1 - \sigma}{\alpha(1 - \sigma s_{j,t,d} - s_{j,t}(1 - \sigma))}.
\]  

(30)

Then I can recover the marginal costs from the estimates of the mark-ups using the formula 

\[
mc_{jt} = -(p_{jt} - mc_{jt}) + p_{jt}.
\]

Table 2 reports some descriptive statistics about the mark-ups and marginal costs. Note that the specification of the demand model implies positive marginal costs for each firm. Then, for each year, I compute an average marginal cost, where the marginal cost of each firm is weighted by the within market share. I do a similar procedure to construct an index for the industry markup. Figure 5 and Figure 6 show the evolutions of these indexes for the industry marginal cost and for the industry markup.

<table>
<thead>
<tr>
<th>Table 2. Descriptive statistics for mark-ups and marginal costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs</td>
</tr>
<tr>
<td>mark-ups ($1982)</td>
</tr>
<tr>
<td>marginal costs ($1982)</td>
</tr>
</tbody>
</table>

Figure 5. Evolution of marginal costs over time (1980 = 100)
I am also able to recover the consumer surplus for each time period using the formula:

\[ CS_t = \frac{1}{\alpha} \ln \left(1 + D_t^{1-\sigma} \right), \]  

(31)

where \( D_t = \sum_{j=1}^{J_t} \exp \left( \frac{\delta_{j,t}}{1-\sigma} \right) \), and \( \delta_{j,t} = x_{j,t} \beta + \theta k_{j,t} - \alpha p_{j,t} + \xi_{j,t} \). Figure 7 reports the evolution of consumer surplus over time (see also Table 3). I find that the consumer surplus is almost multiplied by 10 during the period 1980-2006.
Table 3. Descriptive statistics for the consumer surplus

<table>
<thead>
<tr>
<th></th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumer welfare</td>
<td>27</td>
<td>.0008449</td>
<td>.0005686</td>
<td>.0001137</td>
<td>.0017888</td>
</tr>
</tbody>
</table>

Using the evolution of the freight prices (Figure 9), the evolution of the industry marginal costs (Figure 5), the evolution of the industry’s mark-ups (Figure 6), and the evolution of the consumer surplus over time (Figure 7), I am able to see how the railroad firms have passed the efficiency gains due to takeovers to the final consumers.

From a cost perspective, the industry concentration led to a decrease of the marginal costs (80%) for the period 1980-2004. This is consistent with the literature on cost function (Berndt et al. (1993a, 1993b), Ivaldi and McCullough (2001, 2008)). Moreover, Dennis et al. (2010) mention other efficiency gains of takeovers, leading to a more efficient use of equipment by rerouting traffic to avoid congestion on routes, reduction of ton-miles by choosing the shortest routes, and routes dedicated to specialized services, which might decrease the marginal costs as well.

Moreover, the consumer surplus is almost multiplied by ten during the period 1980-2004 (see Figure 7 and Table 3). This is mainly due to two elements. First, the railroad firms passed part of the decrease in marginal costs to the consumers by decreasing the final prices (60%), and this decrease in the final prices had a positive impact on the consumer surplus. The second element is the increase in the efficiency index of the industry due to the concentration of the industry (2.98 billion $1982 from 1980 to 2004, that is 45.70%, see Figure 8). This means that the takeovers have led to a reallocation of assets from the less efficient firms to the most efficient firms, which has increased the efficiency of the industry. Moreover, this increase in the industry efficiency led to an increase in the consumer welfare during the period 1980-2004. Using the estimates of Table 4, I compute the elasticities of the

---

18 I construct this index by computing a weighted average, where each recovered firm efficiency $\xi_{j,t}$ is weighted by the within market share.

19 Since the recovered efficiency index is negative, it means that the inefficiency has decreased over time.
consumer welfare with respect to the efficiency index and the price index of the industry. I obtain that
an increase of 1% of the industry efficiency implies, on average, an increase of 7.28% of the consumer
welfare and a decrease of 1% of the freight price implies, on average, an increase of the consumer
welfare by 4.66% for the period 1980-2006.\textsuperscript{20} This means that the increase in efficiency, due to the
reallocation of assets from the least efficient to the most efficient firms, had a larger impact on the
consumer welfare than a price decrease due to economies of density. This can be essential for merger
analysis. In general, during a merger investigation, competition authorities mainly focus on the price
aspect. However, this paper emphasizes that the asset reallocation aspect should also be considered
together with the prices, since the former has a significantly large effect on the consumer welfare.\textsuperscript{21}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure8.png}
\caption{Evolution of the industry efficiency index over time}
\end{figure}

\textsuperscript{20} I also find that an increase of 1% of the observed quality index leads to an increase of the consumer
welfare by 2.31%.

\textsuperscript{21} The argument of asset reallocation from least efficient to most efficient firm can also be considered as
an argument in favor of concentration during a merger case.
To summarize, the estimates yield that the takeover waves have led to a decrease in prices due to lower marginal costs and to an increase in the industry efficiency due to reallocation of assets through takeovers. Moreover, the impact of the industry efficiency on the consumer welfare is larger than the impact of prices.

Regarding the mark-ups, Figure 6 shows the evolution of mark-ups over time.\textsuperscript{22} From 1980 to 1999, Figure 6 indicates that the takeovers were profitable: the industry mark-up increased by 20%. However, starting from 1999, I observe that the industry mark-up has stopped to rise. Interestingly, this corresponds to the end of the takeover waves. For instance, in 1999, the Surface Transportation Board prevented a merger between Burlington Northern and Canadian National.

Lastly, from 2004 to 2006, I observe an increase in the final prices by 11.27%. This raised a debate in the US regarding the captivity of shippers and the abuse of market power by the railroad firms. However, I find that the industry marginal cost increased by 26.74% during the same period. This means that the increase in final prices was mainly due to an increase in the marginal costs rather than an abuse of market power.

\textsuperscript{22} The construction of the index for the industry mark-up is similar to the price and marginal cost indexes.
6 Conclusion

In this paper, I analyzed the impact of the takeovers in the US rail freight industry on the marginal costs, prices, and consumer welfare. I find that the concentration led to a decrease in marginal costs by 80%, which was translated into lower prices by 60%. This is coherent with the literature on the estimation of cost function in the US railroad industry, which finds important economies of density. There could also be other efficiency gains of takeovers leading to more efficient use of equipment by rerouting traffic to avoid congestion on routes, reduction of ton-miles by choosing the shortest routes, and routes dedicated to specialized services, which might decrease the marginal costs as well. However, for the period 2004-2006, I find that the industry marginal cost increased by 26.80%, and this was passed to the consumers through a price increase of only 11.3%. This means that the price increase of freight services after 2004 was mainly due to an increase of marginal costs and not to an abuse of market power by the railroad firms.

Moreover, during the period 1980-2006, the concentration of the industry led to a reallocation of assets from the less efficient firms to the more efficient firms. This increase in industry efficiency also positively impacted the consumer welfare. Overall, the decrease in prices and the increase in the industry efficiency led an important increase of the consumer welfare (almost multiplied by ten) during the period 1980-2004. Moreover, this paper shows that focusing on this reallocation aspect can be as important as the price aspect. This can be essential for merger investigations in practice. As a next step, it would be interesting to analyze the data up to today to see if the evolutions of the marginal costs, freight prices, mark-ups, and consumer surplus have the similar trend as in the period 2004-2006.

From a technical point of view, the concentration of the industry led to an unbalanced panel characterized by attrition due to the selection of the most efficient firms. I propose an estimation algorithm to deal with this issue in the context of a structural demand model. The estimation methodology corrects for selection by using a selection equation that is derived from a dynamic model.
that allows for the equilibrium force of exit. I find that the methodology correcting for selection bias
gives more plausible parameter estimates than the standard methodologies. This is an important step
since the estimation of demand model is often one of the most important ingredients in the empirical
industrial organization literature.
APPENDIX 1: THE DATA

The main sources of data are the “Analysis of Class1 Railroads” (hereafter “Analysis”) published annually by the Association of American Railroads (AAR). The “Analysis” is based on regulatory reports that railroads submit to the Surface Transportation Board (STB). In order to adjust for the effect of inflation, I convert the monetary variables in current dollars ($1982) using the Consumer Price Index from the Statistical Abstract of the US (see also the US Bureau of Labor Statistics). I focus on a panel of 42 Class1 rail companies that operated in US between 1980 and 2006. These 42 Class 1 firms are defined as accounting entities (see Table 5 and Table 6).²³

<table>
<thead>
<tr>
<th>Railroad</th>
<th>Years in data</th>
<th>Abbreviation (used in Figure 2 and Figure 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltimore &amp; Ohio (BO)</td>
<td>1978-1985</td>
<td>BO (into CSX in 1985)</td>
</tr>
<tr>
<td>Chesapeake &amp; Ohio (CO)</td>
<td>1978-1985</td>
<td>CO (into CSX in 1985)</td>
</tr>
<tr>
<td>Consolidated Rail Corp. (CR)</td>
<td>1978-1998</td>
<td>CR (splitted between CSX and NS in 1999)</td>
</tr>
<tr>
<td>CSX Transportation (CSX)</td>
<td>1986-2006</td>
<td>CSX</td>
</tr>
<tr>
<td>Norfolk Southern (NS)</td>
<td>1986-2006</td>
<td>NS</td>
</tr>
<tr>
<td>Norfolk &amp; Western (NW)</td>
<td>1978-1985</td>
<td>NW (into NS in 1985)</td>
</tr>
<tr>
<td>Seaboard System Railroad (SBD)</td>
<td>1978-1985</td>
<td>SBD (into CSX in 1985)</td>
</tr>
<tr>
<td>Southern Railway System (SOU)</td>
<td>1978-1985</td>
<td>SOU (into NS in 1985)</td>
</tr>
<tr>
<td>Western Maryland (WM)</td>
<td>1978-1983</td>
<td>WM (into BO in 1983)</td>
</tr>
</tbody>
</table>

²³ Firms that lost their Class1 status during 1978-2006 are excluded from the data. These firms did not really exit from the industry but they disappeared from the Class 1 railroads due to changes in the rule that defines a Class 1 railroad firm. Currently, the Surface Transportation Board (STB) defines a Class I railroad in the United States as having annual carrier operating revenues of $346.8 in 2006.
### Table 6. Railroad firms in the Western area

<table>
<thead>
<tr>
<th>Railroad</th>
<th>Years in data</th>
<th>Abbrevation (used in Figure 2 and Figure 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atchison, Topeka &amp; Santa Fe (ATSF)</td>
<td>1978-1995</td>
<td>ATSF (into with BN in 1995)</td>
</tr>
<tr>
<td>Burlington Northern (BN) ; Burlington Northern Sante Fe (BNSF)</td>
<td>1978-2006</td>
<td>BN ; BNSF</td>
</tr>
<tr>
<td>Canadian National Grand Trunk Corporation (CNGT)</td>
<td>2002-2006</td>
<td>CNGT (it incorporates all US activities of Canadian National Railroad, which included GTW activities)</td>
</tr>
<tr>
<td>Denver, Rio Grande &amp; Western (DRGW)</td>
<td>1978-1993</td>
<td>DRGW (into SP in 1993)</td>
</tr>
<tr>
<td>Detroit, Toledo &amp; Ironton (DTI)</td>
<td>1978-1983</td>
<td>DTI (into GTW in 1983)</td>
</tr>
<tr>
<td>Grand Trunk &amp; Western (GTW)</td>
<td>1978-2001</td>
<td>GTW</td>
</tr>
<tr>
<td>Illinois Central (Gulf) (IC)</td>
<td>1978-1998</td>
<td>IC (into GTW in 1998)</td>
</tr>
<tr>
<td>Kansas City Southern (KCS)</td>
<td>1978-2006</td>
<td>KCS</td>
</tr>
<tr>
<td>Milwaukee Road (MILW)</td>
<td>1978-1984</td>
<td>MILW (into SOO in 1984)</td>
</tr>
<tr>
<td>Missouri-Kansas-Texas (MKT)</td>
<td>1978-1987</td>
<td>MKT (into UP in 1987)</td>
</tr>
<tr>
<td>Missouri Pacific (MP)</td>
<td>1978-1985</td>
<td>MP (into UP in 1985)</td>
</tr>
<tr>
<td>Saint Louis and San Francisco (SLSF)</td>
<td>1978-1979</td>
<td>SLSF (into BN in 1979)</td>
</tr>
<tr>
<td>Saint Louis, Southwestern (SSW)</td>
<td>1978-1989</td>
<td>SSW (into SP in 1989)</td>
</tr>
<tr>
<td>SOO Line (SOO)</td>
<td>1978-2006</td>
<td>SOO</td>
</tr>
<tr>
<td>Southern Pacific (SP)</td>
<td>1978-1996</td>
<td>SP (into UP in 1996)</td>
</tr>
<tr>
<td>Union Pacific (UP) ; Union Pacific-Southern Pacific (UPSP)</td>
<td>1978-2006</td>
<td>UP ; UPSP</td>
</tr>
<tr>
<td>Western Pacific (WP)</td>
<td>1978-1985</td>
<td>WP (into UP in 1985)</td>
</tr>
</tbody>
</table>

Figure 2 and Figure 3 list all the takeovers that happened in the railroad industry (see section 2).

I define a takeover between two firms such that one firm buys another firm. There are two elements of ambiguity for the construction of the merged entities, namely the merged firms CSX and NS in 1986. These two firms appear in 1986 and are the results of the mergers of several firms. I assume that the
merger parties have sold their assets to the firm with the highest market share before the merger. Thus, I assume that the firms BO and CO have sold their assets to SBD in 1986, and the firm NW has sold its assets to SOU in 1986. This treatment of merger yields an unbalanced panel data with an attrition characteristic such that (see Wooldridge, 2002, Chapter 17):

\[ r_{j,t} = 1 \Rightarrow r_{j,r} = 1, \text{ for all } r \leq t - 1. \]

Regarding the construction of the price of providing freight services, I build the series in the ton-miles unit. In particular, for each firm \( j \) active in year \( t \), the “Analysis” gives the Total Gross Freight Revenue (line 599) and the Total Ton-Miles (line 711). I compute the price of freight in ton-miles using the formula:

\[
p_{j,t} = \frac{\text{Total gross freight revenue of firm } j \text{ at year } t}{\text{Total ton-miles of firm } j \text{ at year } t}.
\]

This allows us to build price series that are consistent with the study of the Surface Transportation Board (STB), “Study of Railroad Rates: 1985-2007” (2009), using the data from the “Analysis of Class1 Railroads” (1980-2006). Indeed, the STB has access to confidential and very detailed data (in particular the Official Waybill Sample that records the prices of the commodities shipped in the US), whereas I have access to the “Analysis of Class 1 Railroads” (1980-2006) where pricing information is not directly available. For a particular year \( t \), the industry price index is computed by a weighted average of the prices \( p_{j,t} \) of active firms, where the weights are equal to the market share of firm \( j \) at year \( t \):

\[
s_{j,t,\text{g}} = \frac{\text{Total ton-miles of firm } j \text{ at year } t}{\text{Total industry ton-miles at year } t}.
\]

I compute two industry price indices, namely one price index where the Total Gross Freight Revenue is in current dollars and another price index where it is in $1982 using the Consumer-Price-Index as a deflator (see the Statistical Abstract of the US). In Figure 9, I see that the evolution of the

---

24 This assumption reflects what I observe in the data for all the railroad firms in Figure 2 and Figure 3.
price index is consistent with the evolution of the price index built by the Surface Transportation Board (2009, see Figure 10). Thus, using ton-miles allows to consistently reproduce the evolution of railroad rates displayed in the railroad rates study of the Surface Transportation Board (2009). Secondly, using data from the US Department of Transport (Bureau of Transportation Statistics), I obtain the total size of the freight market in the US (that is the freight provided by air, truck, railroad, water, and pipeline) in the ton-miles unit. Thus, I can construct the market share of each railroad firm (and the market share of the outside alternative) by not making an arbitrary assumption about the total size of the freight market.

**Figure 9. Industry price index (unit of measure: ton-miles, in real $1982, 1980 = 100)**

**Figure 10. Rail rate index (1985 to 2007). Real revenue per ton-miles (1985 = 100)**
The construction of the capital stock follows the methodology of Berndt, Braeutigam, Friedlaender, McCullough, and Meyer (1992). Accordingly, I start from an authoritative estimate of the reproduction cost of capital in 1973 using Nelson (1975) and update the stock of capital of firm $j$ using the perpetual inventory relation:

$$K_{j,t+1} = K_{j,t}(1-d) + I_{j,t},$$

(32)

where $I_{j,t}$ represents the real investment (in $1982$) at year $t$. The depreciation rate $d$ is derived by solving an equation that allows railroad capital to depreciate exponentially over 25 years to a salvage value of 10 percent.\(^{25}\) The “Analysis of Class1 Railroads (1980-2006)” allows the measurement of the nominal investment which is then converted into real value ($1982$). The main difficulty consists in measuring this nominal investment component for way and structures capital. Before 1982, railroads used “betterment” accounting in which the work on railroad way and structures is listed as an expense and thus excluded from the undepreciated book value of road (line 67 in the “Analysis”). Thus a first difference of the undepreciated book value of road allows measuring the nominal investment at every year. After 1982, the railroad industry adopted a depreciation accounting system, where the work on way and structures is added to the book value of road. It is thus necessary to remove the expenditures linked to the maintenance of the network (line 174 minus line 172 in the “Analysis”) from the undepreciated book value of road and then do a first difference to obtain the nominal investment. This perpetual inventory process is iterated to bring the series of way and structure capital until 2006.\(^{26}\)

\(^{25}\) The 25 year assumption is based on Berndt et al. (1992).

\(^{26}\) It is important to mention the treatment of takeovers in the construction of the capital stock. For example, consider the takeover between “UP” and “MKT” in 1987 (see Figure 2). The “Analysis” gives us the data on the capital stock at the end of 1987 for “MKT” and “UP” and the data for the capital stock at the end of 1988 for the merged firm “UP_MKT”. To measure the investment of the merged firm “UP_MKT” in 1988, it is necessary to know its capital stock at the beginning of 1988. However, this data is not available in the “Analysis” (This data exists in the initial R1 reports filled by the railroad firms in 1988 but the R1 reports for the period 1978-1995 are no more available except on microfiche in the library of the Surface Transportation Board in Washington DC. Only the R1 reports for the period 1996-2006 are available on the website of the STB). Thus, I make the arbitrary assumption that the capital stock of the merged firm “UP_MKT” at the beginning of 1988 is equal to the sum of the capital stocks of the merging parties “UP” and “MKT” at the end of 1987.
I follow Vellturo (1989) and Berndt et al. (1993a, 1993b) in using a set of exogenous demand-related variables that can be constructed on a firm-specific basis. These variables — coal consumption (CCON), coal production (CPRO), new car registrations (NEWCAR), state population (SPOP), oil prices (OILP), farm income (FARM), and value of shipment from manufacturing (SHIPMENT) — are measured on a state-by-state basis and then aggregated across states to be railroad specific and to conform to each railroads’ operating territory. These aggregations vary from year to year as some railroad firms exit the industry and some other railroad firms extend their networks by buying the assets of the firms that exit the industry. These variables are based on annual data from the Association of American Railroads, the Department of Transport Statistics, the US Energy Information Administration, the US Department of Agriculture, the US Department of Commerce with the “Annual Survey of Manufacturers” and the different economic censuses, the US Federal Highway Administration, and the Statistical Abstract of the US (see Table 7). The estimation results in section 5 are reported by using only the coal consumption (CCON) variable (this variable appeared to be the most important from a demand side perspective, see also section 2 for a justification). Lastly, I include a quadratic time trend in order to capture any unexplained productivity growth. The strictly exogenous variables include the coal consumption (CCON), the quadratic time trend, the BLP instrument of the coal consumption, the miles of road operated (ROAD), the average length of haul (HAUL), the lag of these six variables, and the lag of the BLP instrument for ROAD. The two variables ROAD and HAUL are used as instruments since they are considered as cost-shifters (see Berndt et al (1993a, 1993b), Ivaldi and McCullough (2001, 2008)). The use of the capital stock as instrument is already mentioned in section 4. Table 8 reports the descriptive statistics.

---

27 The construction of the BLP instruments comes from Berry, Levinsohn, and Pakes (1995). For example, for a railroad firm \( j \), the BLP instrument of the exogenous variable \( x_j \) is \( \sum_{r \neq j} x_r \), where \( r \) denotes a competitor of the railroad firm \( j \).
Table 7. Exogenous variables constructed on a firm specific basis

<table>
<thead>
<tr>
<th>Variable</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coal Production</td>
<td>US Department of Energy, Energy Information Administration, State Energy Data System.</td>
</tr>
<tr>
<td>New Automobile Registration</td>
<td>US Department of Transportation, Federal Highway Administration.</td>
</tr>
<tr>
<td>State Population</td>
<td>US Department of Commerce, Statistical Abstract of The United States, and United States Department of Agriculture</td>
</tr>
<tr>
<td>Oil Price</td>
<td>US Department of Energy, Energy Information Administration, Domestic Crude Oil First Purchase Prices by Area.</td>
</tr>
</tbody>
</table>

Table 8. Descriptive statistics on US Class1 railroad data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unit</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>$1982 Ton-Miles</td>
<td>353</td>
<td>.0260265</td>
<td>.0114441</td>
<td>.0103483</td>
<td>.0853767</td>
</tr>
<tr>
<td>Output $K_{j,t}$</td>
<td>$1982 (000)$</td>
<td>353</td>
<td>9.21e+07</td>
<td>1.26e+08</td>
<td>1285901</td>
<td>6.42e+08</td>
</tr>
<tr>
<td>$k_{j,t} = \ln(K_{j,t})$</td>
<td>$1982$</td>
<td>353</td>
<td>7.604252</td>
<td>1.111563</td>
<td>4.953455</td>
<td>9.368649</td>
</tr>
<tr>
<td>Coal Consumption</td>
<td>Thousand Short Tons (000)</td>
<td>353</td>
<td>265.736</td>
<td>169.3512</td>
<td>11.98112</td>
<td>681.3316</td>
</tr>
<tr>
<td>ROAD</td>
<td>1000 Miles</td>
<td>303</td>
<td>10.34329</td>
<td>9.131503</td>
<td>0.527</td>
<td>35.208</td>
</tr>
<tr>
<td>HAUL</td>
<td>100 Miles</td>
<td>303</td>
<td>5.992068</td>
<td>8.505216</td>
<td>1.75</td>
<td>142.33</td>
</tr>
</tbody>
</table>
APPENDIX 2: ROBUSTNESS CHECK

In the paper, I have derived the correction term for attrition, the Mills ratio, by assuming that the scrap value follows a standard normal distribution. The standard literature on selection with the Heckman two-step procedure has guided the choice of the normal distribution.

As a robustness check, I assume that the scrap value follows an exponential distribution, \( \phi_{j,t} = \exp(1) \). This ensures that the scrap value is strictly positive.\(^{28}\) The threshold value \( \bar{\phi}_{j,t} \) is recovered using the formula \( \bar{\phi}_{j,t} = G^{-1}(\hat{\zeta}_{j,t}) \), where \( \hat{\zeta}_{j,t} \) is the probability of remaining in the industry and \( G^{-1}(.) \) is the inverse of the cumulative distribution function of the exponential distribution. In this case, the correction term for attrition has the following form:

\[
\lambda_{j,t} = 1 - \bar{\phi}_{j,t} \left( \frac{\exp(-\bar{\phi}_{j,t})}{1 - \exp(-\bar{\phi}_{j,t})} \right).
\]

I report the estimates in the case of serial correlation and attrition, that is:

\[
\xi_{j,t} = c_j + \gamma \xi_{j,t-1} + \rho \phi_{j,t-1} + \tau_{j,t}.
\]

Table 9 shows that the parameter estimates with an exponential distribution are very close to Table 1 with a normal distribution for the scrap value. The only difference lies in the standard errors, which are slightly larger in the exponential case.

\(^{28}\) Actually, with a normal distribution, the scrap value is also positive. This can be seen from section 5. Indeed, the recovered thresholds are strictly positive, \( \bar{\phi}_{j,t} > 0 \); since a firm exists the market if the scrap value is higher than the threshold \( \bar{\phi}_{j,t} \), hence the scrap value is also strictly positive.
Table 9. Demand estimates: robustness check

<table>
<thead>
<tr>
<th></th>
<th>Case 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>attrition</td>
<td>Yes</td>
</tr>
<tr>
<td>serial correlation</td>
<td>Yes</td>
</tr>
<tr>
<td>price (−α)</td>
<td>-67.7984*** 22.7286</td>
</tr>
<tr>
<td>σ</td>
<td>0.5423 (0.4093)</td>
</tr>
<tr>
<td>serial corr term: γ</td>
<td>0.1683 (0.2180)</td>
</tr>
<tr>
<td>correction term: ρ</td>
<td>0.5671* (0.2956)</td>
</tr>
<tr>
<td>k_{t,t}</td>
<td>0.2881 (0.2526)</td>
</tr>
<tr>
<td>CCON</td>
<td>0.0006 (0.0015)</td>
</tr>
<tr>
<td>temps</td>
<td>-0.0965 (0.0588)</td>
</tr>
<tr>
<td>temps2</td>
<td>0.0023* (0.0011)</td>
</tr>
<tr>
<td>N</td>
<td>298</td>
</tr>
<tr>
<td>deg freedom</td>
<td>7</td>
</tr>
<tr>
<td>sargan</td>
<td>6.28617</td>
</tr>
<tr>
<td>p value</td>
<td>0.5068</td>
</tr>
</tbody>
</table>
REFERENCES


