Centre for Transport Studies
STOCKHOLM

# ESTIMATION OF ACCESSIBILITY ELASTICITIES IN CONNECTION WITH THE ÖRESUND FIXED LINK USING A PANEL OF MICRO-DATA

Tom Petersen – KTH

CTS Working Paper 2011:10

*Abstract*
The productivity of public infrastructure has been the subject of numerous studies during the last two decades, often with vastly differing results. Matters of concern for these estimates have been the level of aggregation of the data, the measurement of the infrastructure stock, and endogeneity bias. In an attempt to estimate the wider economic impacts of the Öresund fixed link, these issues are addressed by estimating production functions from firm data in Scania – the Swedish part of the Öresund region – using a novel method due to Olley and Pakes (1996), that takes endogenous input choices and self-selection into account. As a measure of the service provided by the infrastructure, accessibility to the workforce is used on a fine-grained geographic level. The sign and significance of the two sources of bias are tested, as well as the robustness of the accessibility parameter with respect to the specification of the barrier of trips across Öresund.

*Keywords*: Firm performance, agglomeration, market potential, accessibility, Öresund region, Olley and Pakes.

*JEL Codes*: R11, R15, H54

Centre for Transport Studies
SE-100 44 Stockholm
Sweden
*www.cts.kth.se*

vti

KTH VETENSKAP OCH KONST

# Estimation of accessibility elasticities in connection with the Öresund fixed link using a panel of micro-data

**Abstract**

The productivity of public infrastructure has been the subject of numerous studies during the last two decades, often with vastly differing results. Matters of concern for these estimates have been the level of aggregation of the data, the measurement of the infrastructure stock, and endogeneity bias. In an attempt to estimate the wider economic impacts of the Öresund fixed link, these issues are addressed by estimating production functions from firm data in Scania—the Swedish part of the Öresund region—using a novel method due to Olley and Pakes (1996), that takes endogenous input choices and self-selection into account. As a measure of the service provided by the infrastructure, accessibility to the workforce is used on a fine-grained geographic level. The sign and significance of the two sources of bias are tested, as well as the robustness of the accessibility parameter with respect to the specification of the barrier of trips across Öresund.
**Keywords:** Firm performance, agglomeration, market potential, accessibility, Öresund region, Olley and Pakes.

## 1 Introduction

The productivity of public infrastructure has been the subject of hundreds of studies since the 1970's with a surge in the 1990's after the publication of Aschauer (1989) and the following debate (Munnell, 1990, 1992; Tatom, 1991, 1993).

The motivations behind this interest are partly to provide guidance for policy decisions, for example as input to cost-benefit analysis of so-called "wider effects", partly to understand the complex role of infrastructure and in the economy, and (more recently) to provide evidence for theoretical models in new economic geography, new growth and new trade theories. The literature on the effect of public infrastructure started as an attempt to explain the productivity slowdown in the

US economy from the 1970's and onwards. For comprehensive reviews of this literature, see for example Gramlich (1994); Sturm, Kuper, and de Haan (1998); Romp and de Haan (2007); Mikelbank and Jackson (2000); Lakshmanan (2010); Straub (2008); Bell and McGuire (1997).

About a decade earlier, as a consequence of a decline in the population in large metropolises, the economies of city size were examined and speculations about the "optimal size" of cities were formed (Mera, 1973). This gave rise to a rich theoretic and empirical literature trying to pin down the foundations and forces of agglomeration economies.

In two recent meta-analyses, the number of studies considered were 67 about the private output elasticity of public capital and 34 about agglomeration externalities (Melo, Graham, and Noland, 2009; Bom and Ligthart, 2009), and these numbers are not exhaustive.

Bom and Ligthart (2009) estimate the short-run output elasticity of agglomeration, corrected for bi-directional publication bias, to 0.085 and the long-run elasticity to 0.27, but with a large heterogeneity among estimates. They report that the main reasons for the heterogeneity are: the empirical model, estimation technique, the type of public capital and the level of aggregation of the public capital data. They find the long-run elasticity to be consistent with the earlier time-series estimates, and that the "primary" (uncorrected) elasticities are significantly inflated by publication bias. Melo et al. (2009) question why one would expect the same elasticities in

Early time-series studies were based on the total US economy or federal state economies, but later on it became more common with smaller geographical divisions: county or metropolitan areas. The econometric problems facing these estimations are manifold, but the most important are perhaps endogeneity of input choices, missing variables, spurious regression, spatial autocorrelation and measurement error. Their respective seriousness depend, among other things, on the spatial and temporal resolution of the data.

Another review of agglomeration, with a focus on policy conclusions, is presented by Gill and Goh (2010).

The construction of the Öresund fixed link provides a new opportunity to estimate the effect of infrastructure on economic performance. The link connects two large markets, Copenhagen, the capital of Denmark with 1.8 million inhabitants (metropolitan area), and Malmö, the third largest city in Sweden with 0.52 million inhabitants (Greater Malmö area; population figures are from Dec 31, 1999) and was opened on July 1st, 2000. In this paper, I present results from the estimation of agglomeration elasticities in production functions with disaggregated panel data (workplace/plant level), in the Swedish part of the Öresund region (Scania) before and after this major change in the infrastructure took place: between 1995 and

2004.

## 1.1 Estimation issues

Disaggregated panel data are increasingly being used for productivity analysis, because of its availability and its potential to gain new insights in the workings of the economy at the lowest level, which is important for efficient policy making. These new data sources do however not necessarily solve the estimation difficulties of earlier studies on aggregate data, and might in fact introduce new ones. Endogeneity is for example as important at the micro level as it is in the aggregate, although of a different kind. With micro data, endogeneity appears because of our ignorance about the internal information set of the management, about the market situation, future plans etc. Moreover, these unobservables are serially correlated, which is a major challenge to the econometrician.

The endogeneity gives rise to a skewed distribution of the errors, which biases the estimates, and serial correlation deteriorates the efficiency of e.g. the OLS estimator and biases the standard errors downwards, potentially leading to spurious rejection of null hypotheses. The endogeneity bias is also known as *transmission bias* after Marschak and Andrews (1944), because the unobservable productivity affects input requirements, which are transmitted to the output. The usual methods to cope with endogeneity and fixed effects is instrumental variables estimation of some form, using transformed data (e.g. by first-differencing, Within transformation[1], or the orthogonal deviations due to Arellano and Bover (1995)). However, differencing in either form enhances measurement error and biases estimates towards 0 (attenuation bias, see Griliches and Hausman (1986, for example)).[2]

Another endogeneity and source of potential misspecification is the fact that managers not only make decisions about how to continue the business, they also decide whether to discontinue it all together. For the econometrician, this is a source of self-selection or attrition bias in the data, if we assume that the decision to discontinue is not taken at random. Rather, this decision is based on performance indicators, accumulated capital, the reservation price of the business etc. Some of these can be captured through e.g. financial statements, while some–notably productivity and "inside information"—are inevitably hidden for the researcher. For overviews of how to handle the endogeneity issue, and all other aspects of productivity estimation on micro data, see Griliches and Mairesse (1995); Eberhardt

[1] Also called "Fixed effects" estimation.

[2] In the generalised method of moments (GMM) framework, it is possible to take both measurement error and endogeneity into account, although efficiency could be an issue together with the problem of finding the right set of instruments, and having time series of sufficient length (Griliches and Hausman, 1986; Arellano and Bond, 1991; Hansen, 1982; Blundell and Bond, 1998, among others).

and Helmers (2010); Syverson (2010); Gandhi, Navarro, and Rivers (2009)

The omission of relevant variables leads to bias in the estimated parameters if there is correlation between the included and the omitted variables, and between the omitted variables and the dependent variable (i.e., they are "relevant") (Greene, 2000, p. 334). Consider the regression

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \epsilon,$$

where all terms are in vector format, and $\mathbf{X}_1$ and $\mathbf{X}_2$ are matrices with $K_1$ and $K_2$ columns, respectively. Now if $\mathbf{y}$ is regressed on $\mathbf{X}_1$ without including $\mathbf{X}_2$, the estimator is

$$\mathbf{b}_1 = \beta_1 + \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{X}_2\beta_2 + \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\epsilon.$$

Taking the expectation, the last term disappears and we get $E[\mathbf{b}_1] = \beta_1 + \mathbf{P}_{1.2}\beta_2$, where $\mathbf{P}_{1.2} = \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{X}_2$ is a $K_1 \times K_2$ matrix with each column being the slopes in a regression of $\mathbf{X}_2$ on $\mathbf{X}_1$. Unless either $\mathbf{X}_2$ and $\mathbf{X}_1$ are uncorrelated, or the true $\beta_2 \equiv 0$, $\mathbf{b}_1$ will be a biased estimate. Furthermore, the direction of the bias is undetermined: it will depend on the combination of the sign of the correlation between the omitted and included variables, and the sign of the elements of $\beta_2$ (i.e. the "true" slopes in a regression of $\mathbf{y}$ where $\mathbf{X}_2$ is included). If $K_2 > 1$, the bias will also depend on the correlation between the different omitted variables. It is also possible that the impacts of the combination of different variables in $\mathbf{X}_2$ and $\beta_2$ cancel out, so that the resulting bias of $\mathbf{b}_1$ is close to 0. This fact might explain why OLS sometimes seems to be a reasonable estimation method, while the background influences to this result are concealed. In Table 5 I present the counteracting effects of endogenous production and selection on the estimates of the agglomeration elasticity.

## 1.2 Aim of the paper and hypotheses

In this paper I want to examine the nature and size of the effects of higher accessibility on production in different industries. I also want to shed some light on the biases at work when these estimates are compared with ordinary least squares (OLS), caused by unobserved heterogeneity and endogeneous selection. I present estimates of production elasticities of accessibility on cross-sectional level, by 2-digit industry, manufacturing/service and for the whole regional economy. As a robustness check, the estimates are compared for three different levels of barrier across Öresund after the establishment of the fixed link. The OLS estimates are compared with estimates by the Olley-Pakes method and the effect of the omitted information on past performance and . Finally, I use estimated firm-specific technical efficiencies, aggregated by the small market area zones used for traffic analysis, and run regressions on 5-year growth to investigate the longer term connection between raised accessibility and this increase in "geographic" productivity (i.e. productivity by zones).

The hypotheses tested are thus:

1. There is no shift in cross-section production stemming from higher accessibility.

2. There is no bias in OLS estimates of accessibility elasticities.

3. There is no long-term effect of higher accessibility on geographic productivity.

The paper is structured as follows: the next section describes various aspects of the included variables, especially accessibility, followed by a section about the three-stage estimation procedure of Olley and Pakes. In this part a special subsection is devoted to the survival model. After that, the results of both the survival and the production regression models are presented, and some results connected to the changed situation in the Öresund region. More detailed results for the goods and service sectors are attached in the Appendix.

## 1.3 Productivity, technical change and technical efficiency

In general, total factor productivity ($TFP$) is measured as the ratio of an index of output $Y$ to an index of inputs $X$, or equivalently the difference between their logarithms, where the indices are products or sums weighted by their respective output or cost shares (Bauer, 1990). For the production frontier function $y^* = f(\mathbf{x}, t)$, the one period Divisia index of $TFP$ in logarithmic form is

$$TFP = \ln y^* - \sum_i s_i \ln x_i,$$

where $s_i$ are the value shares of input (cost shares), i.e.

$$s_i = \frac{w_i x_i}{C} = \frac{w_i x_i}{\sum_i w_i x_i},$$

with factor prices $w_i$ and $y$ and $x_i$ are volumes of output and inputs. In order to represent the case where a firm is not fully efficient, in the meaning that it produces less than possible at a specific level of inputs, we can premultiply $y^*$ with a factor $\lambda = y/y^*$, where $y^*$ is now the *production frontier*, that is the maximum possible amount that can be produced with input vector $\mathbf{x}$ with period $t$ technology ($0 < \lambda \leqslant 1$) (Farrell, 1957). In logarithmic form we get

$$\ln y = \ln \lambda + \ln f(\mathbf{x}, t)$$

Differentiating with respect to $t$ gives

$$\dot{y} = \dot{\lambda} + \sum_i \frac{\partial f}{\partial x_i} \frac{x_i}{f} \cdot \dot{x}_i + \dot{f} = \dot{\lambda} + \sum_i \varepsilon_i \cdot \dot{x}_i + \dot{f}$$

where dot above means growth rate, and $\varepsilon_i$ is the output elasticity of input $i$. If firms are minimising cost, $\varepsilon_i$ can be expressed as $\varepsilon s_i$, with $\varepsilon = \sum_i \varepsilon_i$ the elasticity of scale, by homogeneity of the production function[3]. If production exhibits constant returns to scale, $\varepsilon = 1$, and factors are paid their marginal product, then the technical production parameters $\varepsilon_i$ equal the cost shares $s_i$.

Now $\dot{\lambda}$ is the relative ("percentage") change in technical efficiency ($TE$) and $\dot{f}$ is overall technical change of the industry ($TC$, a time trend), and $\varepsilon_i$ are the estimated Cobb-Douglas output elasticities. We therefore define a suitable performance measure in log-levels, with $TC$ specified as a quadratic time trend:

$$\ln \lambda = \ln y - \sum_i \beta_i \cdot \ln x_i - \beta_t \cdot t - \beta_{tt} \cdot t^2.$$

where $\beta$-s are now parameters to be estimated. We thus assume constant returns to scale. In practice, since we are using a Cobb-Douglas specification, this often yields scale elasticities close to 1 because of the inherent bias towards this value (see Hoch, 1958), so this performance measure will be close to any estimated productivity using this specification.[4]

# 2 Data

## 2.1 Accessibility

The data can be divided in two sources: the data used for accessibility computations and the firm-specific variables. The accessibility calculations are made from population data and calculated travel times between zones in an irregular lattice of 1,345 zones. The geographical coverage is almost the whole Öresund region, namely Scania in Sweden and in Denmark, the entire Zealand plus Møn[5]. The

[3] Cost minimisation gives input prices $w_i = \gamma \frac{\partial f}{\partial x_i}$, with the Lagrange parameter $\gamma$. Multiply with $x_i$ and sum over inputs to get $C = \gamma f \varepsilon$. Now, $\varepsilon s_i = \frac{C}{\gamma f} \frac{w_i x_i}{C} = \frac{\partial f}{\partial x_i} \frac{x_i}{f} = \varepsilon_i$.

[4] Note also that inversion-based estimators like the one of Olley and Pakes "ignore the variation in input mixes and/or measurement errors in inputs" (Gorodnichenko, 2007) and therefore estaimates of returns to scale (RTS) are biased. RTS is however not the main interest of this study.

[5] According to the definition of Statistics Sweden and Statistics Denmark, the Øresund region also includes the islands Lolland, Falster and Bornholm, which are inhabited by 4.3 % of the total population in the region (Dec 31, 2005).

population measure we use is the population in working age, 16–69 years. Some of the data, especially on the Danish side, has been disaggregated from the municipal level onto the finer zonal lattice, in order to take advantage of the travel times. The population data is divided into four classes of last fulfilled education, but this is not used in the present study. The travel times are computed by the SAMPERS travel demand model for work trips by car. Accessibility is computed as a so-called relative Hansen measure, which means that the attraction variable is the share of the population in each zone out of the region total every year. This avoids spurious effects from growth in the total population later in the regressions.This attraction variable is discounted geographically by a declining function of travel time:

$$acc_{rt} = \sum_{j=1}^{N} \frac{POP1669_{jt}}{\sum_{r=1}^{N} POP1669_{rt}} \cdot e^{-0.038 \cdot TTC_{rjt}},$$

where $acc_{rt}$ is accessibility in zone $r$ in year $t$, $POP1669_{xt}$ is the population betweeen 16 and 69 years in zone $x$ in year $t$, and $TTC_{rjt}$ is travel time by car between zones $r$ and $j$ in year $t$. The discounting (impedance) parameter $-0.038$ is chosen so that with a travel time of one hour (expressed in minutes), there is only a 10 % probability that the population makes the trip, compared to the situation with zero travel time. This is also the parameter used in TransCad[6] for work trips. Both accessibility and travel times are indexed by time, because the travel time across the Strait of Öresund changes dramatically in the year 2000, which affects the accessibility in the whole area although the effect is declining towards the periferies. This fact should provide a good foundation for the estimation of the accessibility elasticity, since it is varying in both the cross-section and time dimensions. In reality, only two matrices of travel times are calculated; before and after the fixed link is introduced. In the year 2000, since the opening was exactly in the middle of the year (July 1), a mean of the before- and after-travel times is used in each zone. The year-to-year variability of the accessibility measure is still guaranteed through the variations in population (which, however, are small in comparison). The accessibility variable is coded on each firm/workplace by the geographical coordinates of the firm.

The changes in accessibility by industry are presented in Table 1. The base accessibility ranges from 0.064 to 0.102, with a weighted average of 0.088 (the overall average base accessibility level). The relative increases vary from 11 to 25 % between industries, with an overall average increase of 18 %. These are quite large increases, mainly due to the reduction in travel time to the Danish capital (a small portion depends on the population increase).

[6] TransCad is a Geographic Information System software for Transport modelling applications (GIS-T), see http://www.caliper.com.

| SNI | avg acc > 2000 | avg acc < 2000 | abs diff | diff in % | avg #firms > 2000 | avg #firms < 2000 |
|---|---|---|---|---|---|---|
| 1 | 0.076 | 0.066 | 0.010 | 15 | 803 | 778 |
| 15 | 0.098 | 0.085 | 0.013 | 16 | 258 | 281 |
| 20 | 0.080 | 0.064 | 0.016 | 24 | 166 | 156 |
| 22 | 0.110 | 0.094 | 0.016 | 17 | 435 | 500 |
| 24 | 0.106 | 0.088 | 0.018 | 21 | 117 | 108 |
| 25 | 0.098 | 0.084 | 0.014 | 17 | 144 | 129 |
| 28 | 0.098 | 0.084 | 0.014 | 17 | 541 | 502 |
| 29 | 0.099 | 0.085 | 0.014 | 16 | 346 | 355 |
| 31 | 0.111 | 0.097 | 0.015 | 15 | 90 | 84 |
| 33 | 0.121 | 0.098 | 0.024 | 25 | 166 | 166 |
| 36 | 0.089 | 0.077 | 0.012 | 16 | 161 | 144 |
| 40 | 0.093 | 0.084 | 0.010 | 11 | 83 | 90 |
| 45 | 0.097 | 0.083 | 0.014 | 17 | 2,043 | 1,969 |
| 50 | 0.098 | 0.083 | 0.015 | 18 | 891 | 907 |
| 51 | 0.109 | 0.092 | 0.017 | 19 | 2,667 | 2,693 |
| 52 | 0.104 | 0.087 | 0.016 | 19 | 2,539 | 2,715 |
| 55 | 0.107 | 0.089 | 0.018 | 20 | 938 | 776 |
| 60 | 0.093 | 0.079 | 0.014 | 18 | 981 | 951 |
| 63 | 0.110 | 0.095 | 0.015 | 15 | 329 | 282 |
| 70 | 0.108 | 0.091 | 0.017 | 18 | 816 | 409 |
| 71 | 0.110 | 0.095 | 0.015 | 15 | 170 | 156 |
| 72 | 0.126 | 0.102 | 0.024 | 24 | 611 | 398 |
| 73 | 0.125 | 0.102 | 0.023 | 22 | 133 | 85 |
| 74 | 0.115 | 0.097 | 0.018 | 19 | 3,631 | 3,141 |
| 80 | 0.109 | 0.095 | 0.014 | 14 | 325 | 158 |
| 85 | 0.106 | 0.091 | 0.015 | 17 | 935 | 870 |
| 92 | 0.108 | 0.090 | 0.018 | 20 | 377 | 288 |
| 93 | 0.105 | 0.089 | 0.016 | 19 | 219 | 199 |
| Goods | 0.096 | 0.083 | 0.013 | 16 | 5,351 | 5,262 |
| Service | 0.108 | 0.091 | 0.017 | 19 | 15,561 | 14,029 |
| All | 0.105 | 0.088 | 0.016 | 18 | 20,912 | 19,291 |

Table 1: The average change in accessibility for the included firms in the dataset, by industry, and the approximate number of firms affected.

The assumptions about the travel impedances are crucial for the accessibility measure, and thus for the estimation of production elasticities. The travel times for the passage over the Öresund Strait are fetched from the regional SAMPERS model (ref Beser-Algers), which was calibrated with regard to trips crossing the Strait. Therefore, the link times are not exactly as the actual travel times, but includes some extra time to account for different barrier effects (which account for that travel is lower than "expected"). These extra amounts also accounts for monetary costs, interchange and waiting times, inconveniences etc.

The passing time time it takes for a car trip from central Malmö to central Copenhagen according to the shortest path algorithm is 133 min before the fixed link (over the ferry link Limhamn-Dragør). After the fixed link opened, the travel time is 40 min. The service of the ferries between Helsingborg and Helsingør in the Northern crossing has not changed essentially. The monetary costs to pass the Öresund Strait are assumed to be approximately equal in real terms before and after the introduction of the fixed link. This is not unreasonable, since the pricing scheme is restricted by the governments not to exercise unfair competition towards the Northern part of the Strait[7]. In order to keep these extra impedances as much as possible, we assume a mean waiting time of half the service interval, as is usual in modelling practice, and calculate the extra impedance in MC to $101 - (55 + 60/2/2) = 31$ minutes, which is added to the driving time on the fixed link, approximately 10 minutes coast-to-coast.[8]

## 2.2 Firm-level data

The firm-specific variables are compiled from the financial accounts of all firms except self-employed, and are obtained from Statistics Sweden for the years 1995–2004 (as of Dec 31 each year). The industrial coverage is the primary sector, manufacturing and service industries, except the financial sector (banks and insurance companies). Due to our need for a well-defined geographical location, firms with several workplaces inside Scania are excluded, and only firms with more than 50 % of their activity in terms of number of employees work there are included. This selection rule might bias the results, since firms with more than one plant/workplace ar likely larger than single-unit firms. On the other hand, the number of single-unit firms is massively dominating the number of multi-unit firms: the average number of workplaces per firm in 2004 was 1.087 in Sweden. But worse, this selection rule is active in all time periods, meaning that if a single-unit firm is transformed into

[7]Except the Limhamn–Dragør line, which carried motorised vehicles, there was a shuttle between central Malmö and the Kastrup airport (Flygbåtarna), but since we calculate car travel times this is not included in the model.

[8]In the HH crossing it is $78 - (25 + 60/4/2) = 45.5$ minutes, although it is not used since we assume the same impedance there before and after 2000.

| | coast-to-coast | | | | between city centres | |
|---|---|---|---|---|---|---|
| | | | model | | | |
| | real | approx. freq. per hour | before 2000 | after 2000 | before 2000 | after 2000 |
| Helsingborg–Helsingør | 25 | 4 | 78 | | 85 | |
| est. A | | | | 25 | | 32 |
| est. B | | | | 49 | | 56 |
| est. C | | | | 78 | | 85 |
| | | | | | | |
| Malmö–Copenhagen, via Limhamn–Dragør | 55 | 2 | 101 | – | 133 | |
| via fixed link | 10 | – | | | | |
| est. A | | | | 10 | | 42 |
| est. B | | | | 34 | | 66 |
| est. C | | | | 41 | | 73 |

Table 2: The real and implemented passage times for accessibility calculations, in minutes.

a multi-unit one, for example through merger or acquisition, it drops out of the dataset in that time period; and conversely, if it sells off a unit, it can reappear. However, there is no workaround for this, because financial accounts are made for firms and not workplaces.

On the other hand, groups of companies *are* represented if their members are autonomous, single-plant firms; this is the case for some grocery chains and franchises, for example. However, groups of companies potentially pose other problems to data management and estimation, since they have the possibility to redistribute certain profits and assets between its members. In order to avoid this, I have used performance measures that are independent of such transactions (such as for example earnings before financial entries and balance-sheet allocations).

Variables on production, investment, capital and value-added are transformed into volumes by industry-specific cost indices. Dummies for start-up and closures of plants, change of location, change of activity and change of owner category are coded from the time series of the id-s, location variable, activity code and ownership class (e.g. public, private or foreign) from the unbalanced data set. The dummy for start-up is constructed from the age variable and available for all years. The dummy for closure is constructed from the last observation of each work-place id, which in a way could be misleading since id-s could change for other reasons than closure (for example change of owner or activity). However, it surely indicates that "something happened" in this year, and for example change of activity is already captured through its own dummy. The closure dummy is of course unavailable in the last year of the panel. The change of location is constructed from changes in coordinates that are not caused by missing geographical coding.

The measure of capital is the book value of capital, and not the perpetual inventory used by many other authors. With the book value method, last years capital is depreciated by actual monetary values of depreciation and this years (net) investment is added, while the perpetual inventory method adds lagged capital, depreciated by a fixed percentage, to current year investment. With both methods restrictive assumptions have to be made about the depreciation in the "real" value of the capital stock to the firm: for example, depreciation rate, vintages, utilisation rate, adjustment costs, etc. The book value method is used here for simplicity. Besides, there are empirical results indicating that it might be more appropriate to *ap*preciate the value of capital in the years following an investment, due to gestation lags (Pakes and Griliches, 1984).

In Table 9 in the Appendix, some summary statistics for each 2-digit industry are presented, as well as for the aggregate goods and service sectors and the whole dataset. It is apparent that the industries are quite different in terms of number, output, average employment and investments, and this should always be kept in mind when analysing them all by more or less the same method as we do here. Note that the unit of gross invesment in the right-most column is 1,000 times greater than the one for net investment. It is also shown that a great deal of observations are lost by both lacking location information (no accessibility observation) and the restriction we have on investment: on average around 40 % missing. Apart from them, all firms with more than one unit are excluded because of their locational ambiguity. The size distribution of the firms in the dataset, as compared to the distribution in Sweden, is shown in Table 3. There is quite good agreement of the distributions up to sizes of 100 employees–above that, the sample is underrepresentative. Whether this depends on regional differences or if it is a "true" underrepresentation is not known.

| #employees | Sweden (%) | | Scania dataset (%) |
|---|---|---|---|
| 0 | 75 | | |
| 1-4 | 17 | 68 | 66 |
| 5-9 | 4.0 | 16 | 18 |
| 10-19 | 2.1 | 8.3 | 8.7 |
| 20-49 | 1.2 | 4.8 | 4.7 |
| 50-99 | 0.4 | 1.4 | 1.4 |
| 100-199 | 0.2 | 0.7 | 0.4 |
| 200-499 | 0.1 | 0.4 | 0.2 |
| 500+ | 0.1 | 0.4 | 0.1 |

Table 3: The size distribution of firms in Sweden and in the Scania dataset as of Dec 31, 2004. The two rightmost columns show the distributions of non-self-employed firms, which is the relevant one in this dataset.

For details on the average sizes of firms in the sample compared to the Sweden average, see Table 4. In 2004, the average size of firms in Sweden excluding self-employed was 6.5 employees. For the "goods" industries (SNI 1–45) it was higher, 11.5, and for the service sectors (SNI 50–93 except for the financial and insurance sectorsc: SNI 65–67) it was 5.0. If electricity, gas, and water supply (SNI 40–41; 20.2 employees/firm) and construction (SNI 45; 4.7 employees/firm) are excluded from the industry sectors, the average size in 2004 rises to 18.7 employees/firm in the remaining goods industries. For the whole period 1995–2004, the average size in goods (except electricity generation and construction, i.e. in SNI 1–37) was 21.5, and in services except finance and insurance 5.3. The all together average size in these industries (SNI 1–37, 50–63 and 70–93) was 7.5. (Note that these two

| | 2004 | | | | |
|---|---|---|---|---|---|
| | | | without SNI 40–45 | | |
| | **Sweden** | **Scania sample** | **Sweden** | **Scania sample** | |
| Goods | 11.5 | 11.5 | 18.7 | 14.5 | |
| Service | 5.0 | 6.0 | 5.0 | 6.0 | |
| All | 6.5 | 7.4 | 6.7 | 7.5 | |
| | | | **avg. 1995 – 2004** | | |
| | | | without SNI 40–45 | | final sample (Table 9) |
| | **Sweden** | **Scania sample** | **Sweden** | **Scania sample** | |
| Goods | | 11.9 | 21.5 | 14.9 | 10.1 |
| Service | 5.3 | 5.8 | 5.3 | 5.8 | 5.8 |
| All | | 7.5 | 7.5 | 7.6 | 7.2 |

Table 4: Average sizes of firms in Sweden and in the Scania sample.

samples are not entirely comparable.)

In my sample from Scania, the overall average size for the whole sampling period is 7.2 employees/firm for all firms (including electricity generation and construction), for the "goods" sectors, including electricity generation and construction, 10.1, and for the service sectors except finance and insurance 5.8 (see Table 9 in the Appendix). Thus, only in the goods sectors the average size is slightly different from the Sweden average. However, this could at least partly be attributed to regional differences: most of the largest companies in for example mining, steel production, paper and pulp production and car and truck manufacturing are located outside Scania.

The manufacturing industries in Scania are dominated by companies in the food and packaging industries (Tetra Pak, Alfa Laval, PLM), rubber and chemical industries (Trelleborg, Perstorp, Boliden Kemi), telecommunications (Sony Ericsson) and medical equipment (Gambro). Other large companies in the "goods" sector are E.ON (electricity supply) and Skanska and PEAB (construction).

In Table 10 in the Appendix, some of the dynamic properties of the industries are listed in grand averages: survival from year to year, frequency of starters (entry), relocation, change in activity code (5-digit SNI) and change in ownership *category* (not ownership in itself). The ownership categories are only a few: public, private without and with group association, or foreign ownership. The rate of foreign ownership is also included. It is immediately visible that for example the survival rates are lower, and the entry and relocation rates are higher in the service sector, while the propensity to change activity or ownership category are about the same, although they vary a lot across two-digit industries.

### 2.3 Censored age variable

The age variable is calculated from the start date of the firm, which is censored to the left at Dec 31, 1971. This in turn means that age is censored to the right at between 24 and 33 years, depending on the year. It also means that age is expressed in fractions of years, if it started somewhere in the middle of the first year. The censored proportion varies from 1 to 51 %, depending on the mean age of the industry, and decreases naturally along the period. In general, industries with high entry and exit rates like *Hotels and restaurants* and *Other business activities*, are those with the lowest median age and also the lowest proportion of censored observations. Censoring in the explanatory variables is essentially a missing-value problem, which mainly affects the estimation efficiency, but if age is correlated with the unobserved idiosyncratic effects (i.e. not missing at random), it can also generate bias. This could be mitigated by imputation. However, although there are methods for this in the duration model literature (Pan, 2001; Hsu, Taylor, Murray, and Commenges, 2007), in the case of censoring this is not trivial, because the imputation must be made outside the range of the variable in question according to some hypothetical distribution. Instead I have chosen to exclude the censored observations.

## 3 Estimation

### 3.1 Olley-Pakes estimation

In the estimation approach of Pakes (1994) and Olley and Pakes (1996), both sources of bias are accounted for. Two "helper functions" are estimated—one accounting for the transmission bias and the other for the self-selection bias—together with the parameters of interest, and predictions from them are included in a "bias function" in the last stage. The bias function is approximated as a Taylor series expansion (polynomial series) of the two helper functions.

The framework for estimation of the parameters in the production function is explained in detail in Ackerberg, Benkard, Berry, and Pakes (2007, sect. 2.3).[9] It

[9] It has been extended by Muendler (2007) to allow for negative net investments, and by Levinsohn and Petrin (2003) for intermediate inputs as an alternative to the investment proxy.

provides a method for taking account of persistent unobserved firm heterogeneity, which induces serial correlation, and endogenous (self-)selection. A firm is supposed to enter the market by investing an entry fee in order to test the competitiveness of its specific entrepreneurial idea. Once in the market, in each time period the firm calculates its expected discounted future returns $V(\omega_t, k_t, a_t, \mathbf{D}_t)$, conditional on its efficiency level $\omega_t$, capital endowments $k_t$, age $a_t$ and a vector of market or environmental conditions $\mathbf{D}_t$. These market conditions are common to at least some of the firms in the market and could include for example input prices, output market characteristics, industry structure, technology, tariffs, regulations, weather etc. In this specific application it also includes the accessibility in the zone where the firm is located. The reason for bringing age into the value function is to separate the cohort effect from the selection effect in determining the impact of age on productivity (Ackerberg et al., 2007).

The firm compares this value with a reservation price (sell-off value) $\Phi_t = \Phi(\omega_t, k_t, a_t, \mathbf{D}_t)$, and if it is less than that it takes $\Phi_t$ and exits from the market. It controls the value of its future efficiency level by investments $NetInvVol_{t-1}$, which increases capital deterministically. In terms of productivity, the returns of the investment is stochastic and follows a first-order Markov process:

$$p\left(\omega_{jt+1} \middle| \{\omega_{j\tau}\}_{\tau=0}^{t}, J_{jt}\right) = p\left(\omega_{jt+1} | \omega_{jt}\right)$$

where $J_{jt}$ is the total information set in period $t$, that is all previous values of all variables from the beginning up to time period $t$. This equation states that productivity in $t+1$ only depends on productivity in $t$, regardless of past history (history is assumed to lead to accumulated knowledge and other factors determining productivity, all available in $t$).

The Bellman equation for an incumbent firm becomes

$$V_t(\omega_t, a_t, k_t, \mathbf{D}_t) = \max_{NetInvVol_t, \chi_t} \{\Phi_t, \sup_{NetInvVol_t \geqslant 0} [\pi_t(\omega_t, a_t, k_t, \mathbf{D}_t) - c(NetInvVol_t) + \\ + \beta \cdot E(V_t(\omega_{t+1}, a_{t+1}, k_{t+1}, \mathbf{D}_{t+1}) | J_t)]\} \quad (1)$$

where $\chi_t$ is the discrete control to continue in operation (if $\chi_t = 1$) or to exit from the market and collect the sell-off value $\Phi_t$ ($\chi_t = 0$), and the discount factor is the constant $\beta$. This is the conceptual framework for a firm with state variables $k$ and $\omega$ solving a dynamic programming problem using the control variables $NetInvVol$ and $\chi$; but we will not attempt to solve for the valuation function in (1); instead, we side-step this rather cumbersome problem and focus on the removal of the most important biases mentioned above.[10]

[10] At the same time, there is a continuous ongoing exogenous process of deterioration of the

The firm is assumed to know the level of all (slowly adjusted) state variables (productivity, capital, age and market conditions) in the beginning of each time period, while inputs of intermediate goods and labour are assumed to be adjusted during the whole time period, as a response to exogenous shocks and changes in the environment **D**. These inputs are termed "non-dynamic", since they only affect production in the current time period. The case is different for the investment control: it is known one period in advance, and affects the production in following time periods (and is thus "dynamic"). This also means that we assume that it takes a full time period (one year) to order, receive and install new capital before it can be productive.

Besides this "timing" assumption, there are two other assumptions in this estimation strategy: first, firm performance, unobserved by the researcher but at least partly known in advance by the firm management (because of serial correlation), is assumed to vary positively and monotonically with investment, at least for firms whith strictly positive investment[11]. Second, productivity is assumed to be the only *unobserved* state variable. The whole idea is to account for the "insider information" of the management of the firm by looking at its actions, *viz.* investment decisions. If the management has high expectations about its firms possibilities in the future market, then it invests, and the higher the beliefs, the larger the amount of investment.

The optimisation problem in (1) results in two decision rules, one for each of the controls:

$$\chi_t = \begin{cases} 1 & \text{if } \omega_t \geqslant \underline{\omega}_t \left(\ln kb_t, \ln km_t, age_t, \ln acc_t\right) \\ 0 & \text{otherwise} \end{cases},$$

and

$$NetInvVol_t = NetInvVol\left(\omega_t, \ln kb_t, \ln km_t, age_t, \ln acc_t\right),$$

where $\underline{\omega}_t(\cdot)$ (the lower limit of productivity before exit) and $NetInvVol_t(\cdot)$ are determined by the market equilibrium in period $t$, and depend on all the variables determining that equilibrium, including e.g. input prices, industry structure, etc.

Given the monotonicity (for $NetInvVol > 0$) of the scalar productivity $\omega$, we can invert the investment control function $NetInvVol = NetInvVol(\omega, \mathbf{D})$ and

efficiency and market prospects, making investments necessary to stay in business. There is also a continuous entry of new firms which contribute to the market structure and thus the level of competition at each stage in the development of the business (Ericson and Pakes, 1995).

[11] Although this has in theory been relaxed in the studies by Muendler and Levinsohn and Petrin (2003). However, empirical results from this study do not support monotonicity for negative net investments (i.e. disinvestments greater than or equal to gross investments in one year). Rather, it seems that a high level of productivity implies both high levels of investment and disinvestment, while low productivity implies none of the two.

express it as a function of investment, capital, age (experience) and accessibility (environmental or market condition)(Pakes, 1994, Theorem 1):

$$\omega_t = NetInvVol^{-1} = h\left(NetInvVol_t, \ln kb_t, \ln km_t, age_t, \ln acc_t\right).$$

Following Olley and Pakes, the production function to be estimated takes the Cobb-Douglas form in logarithms, and includes age[12]. The two things added here compared to their specification is the separated capital variables, in buildings and land on the one hand and machinery and equipment on the other; and the "environment variable" accessibility, which is a variable capturing market size and usually varies slowly in time and space:

$$\ln y_{jt} = \beta_0 + \beta_m \ln m_{jt} + \beta_l \ln l_{jt} + \beta_{kb} \ln kb_{jt} + \beta_{km} \ln km_{jt} + \\ + \beta_{age} age_{jt} + \beta_{acc} \ln acc_{jt} + \omega_{jt} + \eta_{jt} \quad (2)$$

where $y$ is production, $m$ is intermediate inputs, $l$ is labour in full-time equivalent workers per year, $kb$ is capital structures (buildings and land), $km$ capital machinery and equipment, and $acc$ is accessibility to population in ages 16-69 years (measured by a relative Hansen measure); $\omega$ is productivity and $\eta$ is the idiosyncratic error, assumed to be i.i.d. with mean zero. The equation is indexed by individual $j$, which will be suppressed below, and year $t$.

Now, as stated above, the capital stock for the coming year of production is assumed to be known in advance, but intermediate inputs and labour are subject to continuous adjustment (no hire and fire costs). The first step is thus to estimate the coefficients for intermediate inputs and labour consistently, with regard taken to the slow adjustments of the capital stock and accessibility. It is important to note that a change in accessibility can occur for several reasons: by exogenous changes, like changes in the different transport systems or changes in the population within reach by these means of transport, or endogenous to the firm, by moving the establishment to another location.

In the first estimation step, the variables for the sluggish and unobserved components of the equation—capital, age, accessibility, and productivity as a function of investment and the three previously mentioned variables—are replaced by a four-degree polynomial in these variables, in order to catch up endogenous input demand stemming from unobserved heterogeneous productivity:

$$\ln y_t = \beta_m \ln m_t + \beta_l \ln l_t + \varphi_t\left(NetInvVol_t, \ln kb_t, \ln km_t, age_t, \ln acc_t\right) + \epsilon_t$$

[12] Although the Cobb-Douglas functional form is not unproblematic, it is simple and very frequently used. The focus of estimation issues in this paper is on biases rather than on functional form, and in this case it could even be advantageous to use a well-known form. For a review of functional forms, see Mishra (2007).

where

$$\varphi_t = \beta_0 + \beta_{kb} \ln kb_t + \beta_{km} \ln km_t + \beta_{age} age_t + \beta_{acc} \ln acc_t \\ + h\left(NetInvVol_t, \ln kb_t, \ln km_t, age_t, \ln acc_t\right). \quad (3)$$

This equation identifies the coefficients for the variable inputs consistently, but not the fixed inputs. Apparently, it is not possible at this stage to separate out the (linear) effect of the state variables on output, from their effect on the investment decision and on the productivity proxy $h$; instead, the linear parameters in $\varphi$ have to be recovered and $h_t\left(NetInvVol_t, \cdot\right)$ estimated by

$$\widehat{\varphi}_t\left(NetInvVol_t, \cdot\right) - \left(\widehat{\beta}_{kb} \ln kb_t + \widehat{\beta}_{km} \ln km_t + \widehat{\beta}_{age} age_t + \widehat{\beta}_{acc} \ln acc_t\right).$$

In the second step, the survival probabilities are calculated.

#### 3.1.1 Survival model

The survival model is estimated by a logit model of survival in the next period, for the years $1995 - 2003$ (the last year is excluded, because we do not yet know if the firm survives or not):

$$P\left(\chi_{jt} = 1\right) = \frac{\exp\left(V_{j,t-1}\right)}{1 + \exp\left(V_{j,t-1}\right)},$$

where $V_t$ is a function of two kinds of capital (buildings and land, and machinery and equipment); several indicators of firm level performance; accessibility to the population in working ages (16–69 years) on both sides of the Öresund strait; age, and time dummies. The performance indicators include value added as a share of total turnover (VAPTO) and per employee (VAPEmpl), size (Labour), average labour cost (EPEmpl), interest of debts (IntDebts), Solidity and indicators of start in the last year (d_start), change of location (d_chgloc) or ownership category (e.g. private, public or foreign; d_chgown). These variables have been chosen among a greater number of variables, where only the ones that were least correlated were kept. Among them, several investment variables were included but they were excluded in the last estimations without loss of performance. Last but not least, it is important to include time dummies to control for exogenous chocks in demand etc., which are not captured in the variables listed above.

Without variable factors (they were removed in the first estimation step), the conditional expectation of (log) output given current inputs, survival and information at $t$ includes the term

$$E\left[\omega_{jt} | kb_{jt}, km_{jt}, age_{jt}, acc_{jt}, \omega_{j,t-1}, \chi_{jt} = 1\right].$$

where $\chi_{jt} = 1$ if and only if $\omega_{jt} \geqslant \underline{\omega}_t (\ln kb_{jt}, \ln km_{jt}, age_{jt}, \ln acc_{jt})$. If the profit function is increasing in capital the value function must also be increasing, and $\underline{\omega}_t (\cdot)$ *decreasing* in capital. If firms are weel endowed with capital, they can expect higher future payoffs and withstand lower $\omega_{jt}$ realisations. This can potentially give rise to a negative bias in the capital coefficients.

### 3.1.2 Non-linear estimation

In the last step of the Olley-Pakes estimation procedure, the predicted values of the "function of endogenous knowledge", $\widehat{h} (NetInvVol_{jt,} \cdot) = \widehat{\varphi}_t (NetInvVol_{jt,} \cdot) - \widehat{\beta}_{kb} \ln kb_{jt} - \widehat{\beta}_{km} \ln km_{jt} - \widehat{\beta}_{age} age_{jt} - \widehat{\beta}_{acc} \ln acc_{jt}$, and the survival probabilities are assembled into a "bias function", which is dependent on investment and some of the production function coefficients; capital, age and accessibility. The rest of this section is an account of their method, with two kinds of capital, and accessibility added to the specification.

To correct for selectivity, we move the variable inputs (intermediates and labor) to the left-hand side in (2), and take expectations conditional on the information in $t-1$ and on survival, i.e. $\chi_{jt} = 1$.

$$
\begin{aligned}
&E [\ln y_t - \beta_m \ln m_{jt} - \beta_l \ln l_{jt} | J_{j,t-1}, \chi_{jt} = 1] = \\
&E [\beta_0 + \beta_{kb} \ln kb_{jt} + \beta_{km} \ln km_j + \beta_{age} age_{jt} + \beta_{acc} \ln acc_{jt} + \omega_{jt} + \eta_{jt} | J_{j,t-1}, \chi_{jt} = 1] \\
&= \beta_0 + \beta_{kb} \ln kb_{jt} + \beta_{km} \ln km_{jt} + + \beta_{age} age_{jt} + \beta_{acc} \ln acc_{jt} + E [\omega_{jt} | J_{j,t-1}, \chi_{jt} = 1] .
\end{aligned}
$$

The second equality follows from that both types of capital, age and accessibility[13] being known in $t-1$, and that $\eta_t$ by definition is uncorrelated with both $J_{t-1}$ and exit at $t$. Developing the last term, we have

$$
\begin{aligned}
E [\omega_{jt} | J_{j,t-1}, \chi_{jt} = 1] &= E [\omega_{jt} | J_{j,t-1}, \omega_{jt} \geqslant \underline{\omega}_t (kb_{jt}, km_{jt}, age_{jt}, acc_{jt})] \\
&= \int_{\underline{\omega}_t}^{\infty} \frac{\omega_{jt} f (\omega_{jt} | \omega_{j,t-1})}{\{1 - F (\underline{\omega}_t | \omega_{j,t-1})\}} d\omega_{jt},
\end{aligned}
\tag{4}
$$

where $F (\underline{\omega}_t | \omega_{j,t-1}) = \int_{-\infty}^{\underline{\omega}_t} f (\mu_{jt} | \omega_{j,t-1}) \, d\mu_{jt}$ (integrated over the cross-section of firms, i.e. over the index $j$) is the value of the conditional distribution function of the productivity levels in the population of firms at $t$, given their productivity in the previous time period, at the lower threshold for exit $\underline{\omega}_t$. The denominator is thus the total probability mass of the continuing firms. Note also that $\underline{\omega}_t$ is a function of the state variables of all firms, being a market outcome of both global

[13] There is a very small perturbation to the accessibility each year, pertaining to the demographic development, but the part pertaining to travel time costs is assumed to be known well in advance. For example, the decision on the Öresund link was taken in the Danish parliament in 1991 and in the Swedish one in 1994.

(e.g. demand or price) factors, and firm-specific outcomes of productivity. The information set $J_{j,t-1}$ is incorporated in the value of the state variables and the value of $\omega_{j,t-1}$, because of the Markov assumption on the development of $\omega_t$; i.e. $\omega_t$ is assumed to depend only on $\omega_{t-1}$.

The conditional expectation in (4) can be expressed as a function $g$ with two indexes, $g\left(\omega_{j,t-1}, \underline{\omega}_t\right)$. While we already have an estimate of $\omega_{j,t-1}$ in $\widehat{h}_{j,t-1}$ above, $\underline{\omega}_t$ has to be estimated from the predicted probabilities from the survival model, $\widehat{P}_{j,t-1}$, together with the estimate of $\omega_{j,t-1}$ once again. First, we have the survival probabilities

$$\begin{aligned}\Pr\left\{\chi_t = 1|\underline{\omega}_t, J_{j,t-1}\right\} &= \Pr\left\{\omega_{jt} \geqslant \underline{\omega}_t|\underline{\omega}_t, \omega_{j,t-1}\right\} \\ &= \wp_{t-1}\left(\underline{\omega}_t, h_{j,t-1}\left(NetInvVol_{j,t-1}, kb_{j,t-1}, km_{j,t-1}, age_{j,t-1}, acc_{j,t-1}\right)\right) \\ &= P_{t-1}.\end{aligned}$$

Now, if the density of $\omega_t$ conditional on $\omega_{t-1}$ is positive around $\underline{\omega}_t$, it is possible to express the productivity threshold as the inverse

$$\begin{aligned}\underline{\omega}_t \approx \wp_t^{-1}\left(P_{jt}, h_{j,t-1}\right) = \wp_{t-1}^{-1}&\left(P_{j,t-1}, \varphi_{j,t-1}\left(NetInvVol_{j,t-1,}\cdot\right) - \beta_{kb}\ln kb_{j,t-1} - \beta_{km}\ln km_{j,t-1}\right. \\ &\left.-\beta_{age}age_{j,t-1} - \beta_{acc}\ln acc_{j,t-1}\right)\end{aligned}$$

Inserting this into $g\left(\omega_{j,t-1}, \underline{\omega}_t\right)$ gives

$$\begin{aligned}g&\left[\varphi_{j,t-1}\left(NetInvVol_{j,t-1,}\cdot\right) - \beta^{(ka)}X^{(ka)}_{j,t-1},\right. \\ &\qquad\left.\wp_{t-1}^{-1}\left\{P_{j,t-1}, \varphi_{j,t-1}\left(NetInvVol_{j,t-1,}\cdot\right) - \beta^{(ka)}X^{(ka)}_{j,t-1}\right\}\right] = \\ &\qquad\qquad = g\left[P_{j,t-1}, \varphi_{j,t-1}\left(NetInvVol_{j,t-1,}\cdot\right) - \beta^{(ka)}X^{(ka)}_{j,t-1}\right] \quad (5)\end{aligned}$$

where $\beta^{(ka)} = \left(\beta_{kb}, \beta_{km}, \beta_{age}, \beta_{acc}\right)$ and

$$X^{(ka)}_{j,t-1} = \left(\ln kb_{j,t-1}, \ln km_{j,t-1}, age_{j,t-1}, \ln acc_{j,t-1}\right).$$

This leads to an estimating equation that is non-linear in the capital, age and accessibility parameters:

$$\begin{aligned}\ln y_t - \widehat{\beta}_m\ln m_t - \widehat{\beta}_l\ln l_t &= \beta_0 + \beta_{kb}\ln kb_t + \beta_{km}\ln km_t + \beta_{age}age_t + \beta_{acc}\ln acc_t + \\ &+ g\left(\widehat{P}_{t-1}, \widehat{\varphi}_{t-1} - \beta_{kb}\ln kb_{t-1} - \beta_{km}\ln km_{t-1} - \beta_{age}age_{t-1} - \beta_{acc}\ln acc_{t-1}\right) \\ &+ \xi_t + \eta_t \quad (6)\end{aligned}$$

where $g(\cdot)$ has mean $E\left[\omega_t|\omega_{t-1}, \chi_t = 1\right]$ by construction, and thus $\xi_t = \omega_t - E\left[\omega_t|\omega_{t-1}, \chi_t = 1\right]$ has mean zero. $\xi_t$ is the part of the productivity that was unanticipated by the firm in period $t-1$. Note here that the parameters for the state variables appear both linearly in the estimating equation, and inside the $g$ function (in front of the lagged state variables).

# 4 Results

## 4.1 Survival models

The estimated significant parameters from the survival model are presented in Table 11 and 12 as elasticities, evaluated at average values of the regressor in question, and marginal effects, respectively. The difference between the tables is that for the variables in the latter, it is more natural to think of the effect (in percent) of an additional year in the case of age, and in the other cases, the incidence of change of location, change of ownership or start-up in the previous year, respectively. In the first table, the entries are the effect in percent of a 1 percent change in the variable in question.[14]

It can immediately be concluded that the capital stock variables are not the most important ones for explaining the survival of firms. Instead, most explanatory power is derived from a productivity measure (value added per worker), age (also a proxy for experience) and labour cost (average earnings per employee) in 16–18 out of 28 industries), followed by solidity (the quotient of adjusted own capital to total capital), capital turnover rate (the quotient of turnover to total capital) and accessibility to the working population (in 10 out of 28 industries); the quotient of value added to turnover (VAPTO), and size (Labour) in eight; change of location and start-up in the previous yer increase risk in seven and six industries, respectively; equipment or machinery capital and interest of debts are significant in only four, buildings and land capital in two, and last change of ownership category in only one industry. This is in contrast with the original model in Olley and Pakes (1996), where the survival model index function is a polynomial of investment, capital and age. With my richer specification, the capital and investment variables are almost superfluous, especially in the aggregate regressions (Goods, Service and All). Out of their variable set, only age is significant here.

Value added per employee is positively associated with survival, while average salary negatively (except in two industries: Renting of equipment and Research and development). However, the positive effect of the former always outweigh the negative effect of the latter, with a factor of 3–4 in general (with outliers of over 6 for Retail trade and below 2 for Agriculture).

About the main variables of interest, accessibility and age, the former almost always increases the risk of quitting the market among the ten industries where this elasticity is significant. This is a clear indication of the increased competitive pressure in larger markets in these industries. The only exception is Manufacturing

[14] The formula for the elasticity with respect to regressor $i$ is $\beta_i \bar{x}_i \left(1 - \hat{P}(\bar{\mathbf{x}})\right)$, and the marginal effect of regressor $j$ is $\beta_j \left(1 - \hat{P}(\bar{\mathbf{x}})\right)$, where $\mathbf{x}$ is the vector of regressors and a bar above means average value.

of precision instruments, which seems to benefit from increased accessibility in terms of survival, and by a fairly large amount.

In the case of age, the survival probability instead increases in almost two thirds of the industries, by at most 1 % per additional year of age (Hotels and restaurants and Research and development). Slightly smaller, around 0.8 % per additional year, is the effect in Manufacture of furniture and Computer activities. Education, Recreational activities and Other service activities have a marginal effect of around 0.5 % per additional year. In most other industries, like Publishing and printing, Manufacturing of metal products, Construction, Wholesale and Retail trade, Land transport, Real estate and Other business activities, the age effect is around 0.3 % of increase in survival probability per additional year.

## 4.2 Production functions

The general conclusions from the estimations indicate that the Olley-Pakes (OP) estimates of both intermediate inputs and labour in general are quite close to the OLS counterparts, with a few exceptions where they are closer to the Within estimates, or somewhere in-between[15]. It is also evident that the larger the dataset, the smaller the bias. The Within estimation bias is positive for intermediate inputs and negative for labour, in general, but again, in a few industries the intermediate input coefficient is also lower than the OLS estimate. For the rest of the variables, age and accessibility are in general positive or insignificant by OLS, while by Within they can be significant in either direction, positive or negative.

In section A.3.3 in the Appendix, the full tables for the aggregated Goods and Service sectors and the aggregated results for All sectors are presented; however in these tables the number of observations is so high that the OLS biases are attenuated, at least for intermediate inputs and labour. In contrast with Olley and Pakes (1996), I find that the capital coefficients are in general lower with their method than with OLS, i.e. the hypothesised negative bias from overrepresentative exit of firms with smaller capital stock is not supported by the results. The reason for this could be the counteracting effect of “management bias”, i.e. a positive correlation between capital stock and excluded managerial input (which is assumed to be positively correlated with productivity). With Griliches words, “firms with a higher level of entrepreneurial and managerial inputs may be less subject to capital rationing” (1957). This independence between exit and capital stock is also confirmed by the survival model, where hardly any of the industries had significant capital coefficients, given our other set of explanatory variables.

[15] The full results of all 28 industries—OLS, Within and Olley-Pakes, together with panel Durbin-Watson serial correlation and Wooldridge heterogeneity statistics are available from the author on request.

In *Table 5* we can get a view of the bias of OLS compared to OLS with included controls for endogeneous input and exit choice, and with OP estimation.

| | OLS | OLS with $\hat{\varphi}_{t-1}$ and $\hat{P}_{t-1}$ | | | OP |
|---|---|---|---|---|---|
| SNI | $\hat{\beta}_{acc}$ | $\hat{\beta}_{\hat{\varphi}}$ (+) | $\hat{\beta}_{\hat{P}}$ (−) | $\hat{\beta}^{+}_{acc}$ | $\hat{\beta}^{OP}_{acc}$ |
| 1 | 0.004 | 0.50*** | 1.03*** | 0.013 | 0.019 · |
| 15 | 0.010 | 0.89*** | −0.30* | −0.006 | −0.008 |
| 20 | 0.009 | 0.81*** | −0.04 | 0.004 | 0.010 |
| 22 | 0.052*** | 0.70*** | 1.33*** | 0.047** | 0.079*** |
| 24 | −0.046 | 0.57*** | −0.88** | 0.034 | 0.052* |
| 25 | 0.075 · | 0.69*** | 0.65 · | 0.048 | 0.082** |
| 28 | −0.008 | 0.52*** | 1.28*** | 0.006 | 0.009 |
| 29 | 0.001 | 0.50*** | 0.95*** | 0.003 | 0.010 |
| 31 | 0.013 | 0.70*** | 0.83** | −0.009 | −0.053 · |
| 33 | 0.064 · | 1.09*** | −0.18 | 0.057 | 0.083 · |
| 36 | 0.000 | 0.72*** | 0.07 | 0.029 | 0.050* |
| 40 | 0.207** | 0.59*** | −0.25 | 0.066 | 0.066 |
| 45 | 0.020*** | 0.73*** | −0.06 | 0.015* | 0.025*** |
| 50 | 0.017** | 0.55*** | 0.00 | 0.012 | 0.021* |
| 51 | 0.005 | 0.66*** | −0.03 | 0.017* | 0.029*** |
| 52 | 0.017*** | 0.67*** | 0.02 | 0.009 · | 0.015** |
| 55 | 0.009 | 0.96*** | −0.01 | −0.002 | −0.002 |
| 60 | −0.005 | 0.49*** | −0.12 | 0.005 | 0.008 |
| 63 | 0.051* | 0.71*** | −0.02 | 0.025 | 0.038 |
| 70 | 0.007 | 0.73*** | −0.08 | 0.015 | 0.019 |
| 71 | 0.032 | 0.62*** | 0.02 | 0.037 | 0.055 |
| 72 | −0.023 | 0.54*** | −0.11 | 0.002 | 0.009 |
| 73 | 0.004 | 0.97*** | 0.20 | 0.192 | 0.283 |
| 74 | −0.017* | 0.63*** | 0.00 | −0.014 | −0.021* |
| 80 | −0.026 | 0.65*** | −0.08 | −0.042 | −0.077* |
| 85 | 0.037*** | 0.55*** | −0.19 | 0.006 | −0.001 |
| 92 | 0.021 | 0.87*** | −0.22 | 0.036 | 0.065 · |
| 93 | −0.038* | 0.83*** | 0.02 | −0.024 | −0.039* |

Table 5: Bias of OLS production function estimates of the accessibility parameter $\hat{\beta}_{lacc}$ in relation to OLS with the inclusion of lagged productivity ($\hat{\varphi}$) and lagged survival probability ($\hat{P}$). The difference between the fifth and second column constitutes the OLS bias. The value and sign of the $\hat{\varphi}_{t-1}$ $\hat{P}_{t-1}$, together with the correlation between these variables and the accessibility, determine the size and direction of the bias. A positive $\beta_{\hat{\varphi}}$ in general has a positive influence on $\hat{\beta}_{lacc}$ because of the positive correlation between productivity (estimated by $\hat{\varphi}$) and accessibility, while the opposite is true for $\beta_{\hat{P}}$. *** - significant on 0.1 %, ** - 1 %, * - 5 %, . (dot) - 10 %.

### 4.2.1 Tests

I perform three specification tests: one focusing on the first estimation step in (3), and two on the last step in (6). The first one, suggested in Ackerberg et al. (2007), tests the validity of the assumption that the variable inputs ($\ln m_t$, $\ln l_t$) are in fact *variable*, in the sense that they are decided *after* the realization of $\omega_t$, and thus uncorrelated with the idiosyncratic error $\epsilon_t$.This is done by comparing the regression

$$\ln y_t = \beta_m \ln m_t + \beta_l \ln l_t + \varphi_t^0 \left( NetInvVol_t, \ln kb_t, \ln km_t, age_t, \ln acc_t \right) + \epsilon_t^0$$

with

$$\ln y_t = \varphi_t^1 \left( \ln m_t, \ln l_t, NetInvVol_t, \ln kb_t, \ln km_t, age_t, \ln acc_t \right) + \epsilon_t^1,$$

where $\ln m_t$ and $\ln l_t$ are included with higher order terms and interactions in the proxy function $\varphi_t^1$. The resulting error terms $\epsilon_t^0$ and $\epsilon_t^1$ are then compared using ANOVA. If they are not significantly different from each other, then the intermediate inputs and labour are in fact chosen independently of $\epsilon_t^0$, given $\varphi_t^0$ as a proxy for individual heterogeneity and productivity (in period $t$). If they do differ, then there is residual correlation between these inputs and the error term in the first equation, and there is either a dynamic effect of earlier choices of these inputs, or these inputs are not entirely variable conditional on $\varphi_t^0$ (not adjusting to current levels of predetermined inputs)[16]. And in fact, this test rejects the null of equality in all industries, both when only labour is included in $\varphi_t^1$ and when both inputs are included, which suggests that both of these inputs are either not variable, or they are dynamic.

The second and third tests are due to Olley and Pakes and test the validity of their approach by including lagged inputs in the third-stage estimating equation. First they include the presumed variable inputs, labour and intermediate inputs:

$$\begin{aligned} \ln y_t - \widehat{\beta}_m \ln m_t - \widehat{\beta}_l \ln l_t = \beta_{kb} \ln kb_t + \beta_{km} \ln km_t + \beta_{age} age_t + \beta_{acc} \ln acc_t + \\ + g\left( \widehat{P}_{t-1}, \widehat{\varphi}_{t-1} - \beta_{kb} \ln kb_{t-1} - \beta_{km} \ln km_{t-1} - \beta_{age} age_{t-1} - \beta_{acc} \ln acc_{t-1} \right) + \\ + \gamma_m \ln m_{t-1} + \gamma_l \ln l_{t-1} + \xi_t + \eta_t \quad (7) \end{aligned}$$

and second, they include the predetermined capital inputs and productivity shifter age. Here I also include accessibility as a predetermined productivity shifter:

[16] Another possibility is of course that $\varphi_t^0$ is wrongly specified and does not represent $\omega_t$ sufficiently close. Note the difference between *dynamic* and *variable*: dynamic means that the inputs are dependent on earlier choices (error terms in previous periods), while variable means that they are adjusted in response to performance $\omega_t$.

$$
\begin{aligned}
\ln y_t - \widehat{\beta}_m \ln m_t - \widehat{\beta}_l \ln l_t = \beta_{kb} \ln kb_t + \beta_{km} \ln km_t + \beta_{age} age_t + \beta_{acc} \ln acc_t + \\
+ g\left(\widehat{P}_{t-1}, \widehat{\varphi}_{t-1} - \beta_{kb} \ln kb_{t-1} - \beta_{km} \ln km_{t-1} - \beta_{age} age_{t-1} - \beta_{acc} \ln acc_{t-1}\right) + \\
+ \gamma_{kb} \ln kb_{t-1} + \gamma_{km} \ln km_{t-1} + \gamma_{age} age_{t-1} + \gamma_{acc} \ln acc_{t-1} + \xi_t + \eta_t. \quad (8)
\end{aligned}
$$

If the inputs $\ln m_t$ and labour $\ln l_t$ are in fact variable, and if $\widehat{\beta}_m$ and $\widehat{\beta}_l$ are correctly estimated in the first stage, then the left hand side of the first test (7) should be uncorrelated with the lagged values of intermediates and labour, $\ln m_{t-1}$ and $\ln l_{t-1}$. The same should hold if they are static, i.e. they only affect current output (in period $t$), and not output in later periods. Thus if this test fails, at least one of these two assumptions is wrong.

In the case of the second test (8), if the $g$ function correctly transmits the effects of the past productivity $\omega_{t-1}$ through the inverted investment control proxy function in (5) and the current productivity threshold $\underline{\omega}_t$, then there should be little variation left for past levels of the state variables (capital, age and accessibility). However, the power of these tests have been questioned (Ackerberg, Caves, and Frazer, 2006). Especially in the latter test, the risk of multicollinearity between both the past and current levels of the state variables, and between the past levels inside and outside the non-parametric function $g$, is high. Accessibility, as well as the two types of capital, are inherently persistent.

The test for lagged intermediate inputs ($lm$) and labour ($ll$) for the whole dataset is shown in Table 6. The parameters of the lagged variables are significantly separate from zero, but their values are small compared to the parameters of the original specification, which are also not greatly affected.

The result of the second test, for lagged state variables, is found in Table 7, again for the whole dataset. Here the pattern is different. The parameters of the lagged state variables are highly significant. In all cases, the sum of the past and current parameter is approximately equal to the original estimate, which is an obvious sign of the multicollinearity mentioned above. Of course, the state variables are as such very serially correlated, and this result might be expected. This result casts doubt on the relevance of this method for the estimation of accessibility elasticities. It is also possible that adjustments of production to changing accessibility takes longer than one year, which is the forward-looking time span used here.

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| $lm$ | 0.6415 [a] | 0.0010 | 632.69 | 0.000 |
| $ll$ | 0.3156 [a] | 0.0017 | 186.23 | 0.000 |
| $lkb$ | 0.0028 | 0.0003 | 9.21 | 0.000 |
| $lkm$ | 0.0137 | 0.0010 | 14.16 | 0.000 |
| $age$ | -0.0172 | 0.0003 | -56.68 | 0.000 |
| $lacc$ | 0.1045 | 0.0027 | 38.11 | 0.000 |
| $tt$ | 0.0035 | 0.0022 | 1.57 | 0.117 |
| $tt^2$ | -0.0002 | 0.0002 | -0.72 | 0.472 |
| $\gamma_{lm}$ | -0.0042 | 0.0011 | -3.93 | 0.000 |
| $\gamma_{ll}$ | -0.0191 | 0.0018 | -10.63 | 0.000 |

[a] The estimates for $lm$ and $ll$ are the same as in Table 19 (first stage of estimation).

Table 6: Test for significance of lagged values of $\ln m$ and $\ln l$ ($\gamma_{lm}$ and $\gamma_{ll}$). Dependent variable: $ly - \widehat{\beta}_m \ln m - \widehat{\beta}_l \ln l$. All industries.

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| $lm$ | 0.6415 [a] | 0.0010 | 632.69 | 0.000 |
| $ll$ | 0.3156 [a] | 0.0017 | 186.23 | 0.000 |
| $lkb$ | 0.0005 | 0.0005 | 1.06 | 0.289 |
| $lkm$ | 0.0154 | 0.0013 | 11.96 | 0.000 |
| $age$ | -0.0087 | 0.0016 | -5.60 | 0.000 |
| $lacc$ | -0.0343 | 0.0157 | -2.18 | 0.029 |
| $tt$ | 0.0074 | 0.0023 | 3.25 | 0.001 |
| $tt^2$ | -0.0005 | 0.0002 | -2.48 | 0.013 |
| $\gamma_{lkb}$ | 0.0029 | 0.0005 | 6.27 | 0.000 |
| $\gamma_{lkm}$ | -0.0113 | 0.0014 | -8.34 | 0.000 |
| $\gamma_{age}$ | -0.0085 | 0.0016 | -5.33 | 0.000 |
| $\gamma_{lacc}$ | 0.1372 | 0.0161 | 8.52 | 0.000 |

[a] The estimates for $lm$ and $ll$ are the same as in Table 19 (first stage of estimation).

Table 7: Test for significance of lagged values of $lkb$, $lkm$, $age$ and $lacc$ ($\gamma_{lkb}$ etc.). Dependent variable: $ly - \widehat{\beta}_m lm - \widehat{\beta}_l ll$. All industries.

### 4.3 Robustness of estimates

In order to assess the robustness of the production elasticities with respect to accessibility, three different travel impedance barriers over the Strait have been used, resulting in three different accessibility changes between the years before 2000 and the years after (see Table 2). The results are presented in Table 8. It shows that the level of the assumed barrier matters a lot for the estimates.

| **SNI** | **A - no extra barrier** | **B - 24 min uniform HH and MC** | **C - original barrier** |
|---|---|---|---|
| 1 | 0.019* | 0.018· | 0.019· |
| 15 | −0.016 | −0.010 | −0.008 |
| 20 | 0.009 | 0.011 | 0.010 |
| 22 | 0.049*** | 0.077*** | 0.079*** |
| 24 | 0.019 | 0.050* | 0.052* |
| 25 | 0.048· | 0.082** | 0.082** |
| 28 | 0.004 | 0.009 | 0.009 |
| 29 | −0.003 | 0.007 | 0.010 |
| 31 | −0.039 | −0.058* | −0.053· |
| 33 | 0.098** | 0.104** | 0.083· |
| 36 | 0.034* | 0.034* | 0.050* |
| 40 | 0.107** | 0.078* | 0.066 |
| 45 | 0.041*** | 0.058*** | 0.025*** |
| 50 | 0.032*** | 0.037*** | 0.021* |
| 51 | 0.028*** | 0.043*** | 0.029*** |
| 52 | 0.031*** | 0.039*** | 0.015** |
| 55 | 0.020*** | 0.020** | −0.002 |
| 60 | 0.024*** | 0.042*** | 0.008 |
| 63 | 0.088*** | 0.082*** | 0.038 |
| 70 | 0.029 | 0.036· | 0.019 |
| 71 | 0.015 | 0.037 | 0.055 |
| 72 | 0.030 | 0.027 | 0.009 |
| 73 | 0.115 | 0.130 | 0.283 |
| 74 | 0.023** | 0.048*** | −0.021* |
| 80 | 0.044· | 0.049· | −0.077* |

Robustness of accessibility estimates with regard to specification of the barrier reduction of the fixed link. Boldface means significant on the 5 % level. *Continued on next page.*

| SNI | A - no extra barrier | B - 24 min uniform HH and MC | C - original barrier |
|---|---|---|---|
| 85 | 0.034** | 0.066*** | −0.001 |
| 92 | −0.011 | 0.004 | 0.065· |
| 93 | −0.001 | −0.005 | −0.039* |
| Goods | 0.0233*** | | 0.030*** |
| Service | 0.019*** | | 0.092*** |
| All | 0.0233*** | | 0.096*** |

Table 8: Accessibility estimates with different specifications of the barrier after the introduction of the fixed link: cases A, B and C (see Table 2). HH = Helsingborg–Helsingør, MC = Malmö–Copenhagen. Boldface means significant on the 5 % level.

## 4.4 Evaluation in the time dimension

In order to assess the relationship between the performance measure and accessibility in the time dimension, the above OP regressions were repeated on all the data (pooled industries), but now without the accessibility variable. The residuals from this regression was used as a technical efficiency performance measure, which was aggregated zonewise and differenced with 5-year intervals[17]. These differences were then regressed against 5-year differences of accessibility plus year dummies, and plotted together with the data, see

From this plot and the t-value of the slope coefficient, we cannot reject hypothesis that there is no long-term effect of higher accessibility on geographic productivity (hypothesis 3 above), at least not in a five year period. Admittedly, five years is not very long in the perspective of the life span of an infrastructure like this—therefore this results begs for continued studies on longer panel data sets.

[17]The aggregation of the residuals were weighted by the output share. Olley and Pakes (1996) erroneously use the *exponential* of the residual as a productivity measure for post-analysis, which has a great impact because of the presence of extreme outliers. In this case, the productivity also has to be aggregated *geometrically*, not *arithmetically*.

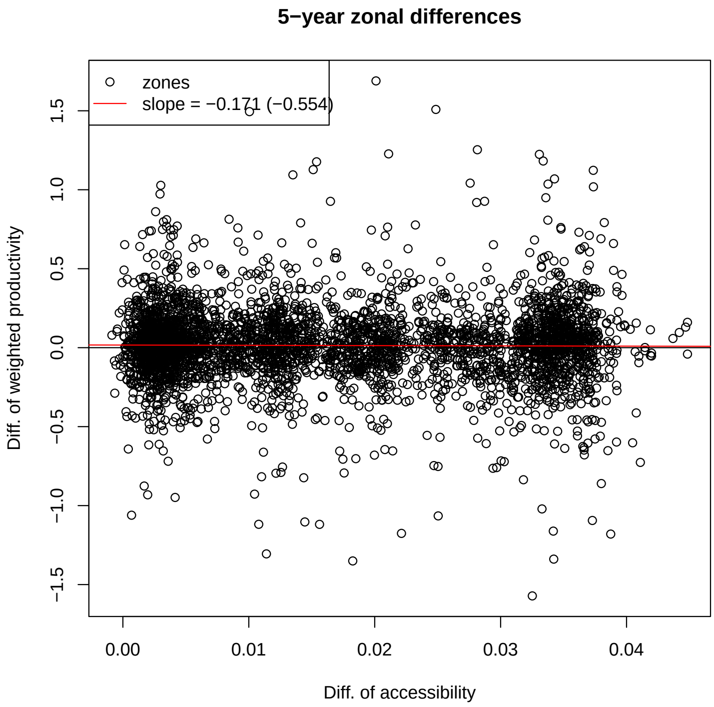


Figure 1: Cross-plot of 5-year differences in aggregated zonal productivity versus 5-year differences in accessibility, 1996-2004. Regression line in red and the regression parameter and t-value in the legend. Year dummies were included in the regression.

# 5 Conclusions

In an attempt to estimate the wider economic impacts of the Öresund fixed link, the issues of infrastructure measurement and endogeneity have been addressed by estimating production functions from firm data in the Swedish part of the Öresund region, using the method of Olley and Pakes (1996). As a measure of the service provided by the infrastructure, accessibility to the workforce is used on a fine-grained geographic level. The sign and significance of the two sources of endogeneity bias have been tested, as well as the robustness of the accessibility parameter with respect to the specification of the barrier of trips across Öresund. Furthermore, consistent estimates of technical efficiency have been extracted, and the before-and-after effect of the fixed link is estimated by a regression in 5-year differences.

It is concluded that a) survival is significantly negatively associated with accessibility in 10 out of the 28 industries tested, b) endogenous input choices constitutes a significant source of positive bias of the accessibility parameter compared to OLS estimates; c) this bias is however partly offset by the bias of the survival probability of firms, d) the construction and trade sectors have a relatively robust higher productivity in higher-accessibility locations, as well as the pooled goods and service samples, and e) until 2004, there is no evidence of increased aggregate productivity from the introduction of the fixed link in July 2000.

Reviewing the three hypotheses set up in subsection 1.2, we can conclude that the first two—"no shift in production stemming from higher accessibility", and "no bias in OLS estimates"—are rejected, but not the third—"no long-term effect of higher accessibility".

In any case, it is apparent from this study that with micro-level data, it is necessary to account for at least input endogeneity, and, in the second place, exit probability, in order to capture the effects of accessibility on output. In earlier studies on data from 1990–98 and using OLS, GLS, Fixed effects estimation on cost functions, and in another study using propensity score matching, also did not exhibit any accessibility effects (Petersen, 2004b,a). Estimations (not reported here) with the Blundell-Bond GMM-SYS estimator (1998) of production functions, using the same dataset as here but without exit probabilities, have given the same result. However, in those cases the Sargan test revealed non-stationarity in some industries—construction, trade and business services—i.e. the same industries that stand out in the results presented here, and which are also expected to be affected by improvements in accessibility.

# 6 Discussion

Part of the lack of time series correlation above could perhaps be explained by the fact that the demand side is excluded from the analysis. If markets are reasonably competitive during this time period, firms' profits will not grow, and productivity enhancements should successively be transfered to the consumers in the form of cheaper, better and more diversified products. This is not reflected in these data on production. Gains in non-monetary benefits like public amenities, better living standard, higher quality products, etc. are not included in the book-keeping of private firms, but they still constitute a significant portion of the total benefits. For example, Mamuneas and Nadiri (2006), in a general equilibrium framework, estimate the benefits accruing to consumers to be 40–55 % of the total rates of return to highway capital in the United States from 1949 to 2000. Skytesvall and Hagen (2006) estimate this proportion in Sweden to be even larger, about 67 % for the period 1993–2003.

Although these results are not consistent with theory, they are not unique: several previous cross-country studies have yielded the same results, also with longer time-series than the one used here (Canning and Fay (1993); Canning and Bennathan (2000, see e.g.). On the other hand, in a shorter and smaller dataset, Åkerman (2009) finds evidence of aggregate productivity gains in Malmö municipality, compared to the other two metropolises in Sweden, Gothenburg and Stockholm, due to the fixed link. These productivity gains stem from increased exports, and are mainly driven by exit of the least productive firms, and the expansion in output of firms with already high productivity, that enter the export market. He finds that this evidence confirms the firm-selection patterns predicted in trade models like Melitz (2003)

It could also be that the Olley-Pakes method in this case fails to take the full account of firm dynamics, which is indicated by the failure of the specification tests–especially regarding the capital and accessibility coefficients. Even if exit probabilities are accounted for, other possible adjustments that affect accessibility are possible, namely relocation. In relation to the large increase in accessibility stemming from the opening of the fixed link, it is probable that gestation lags are longer than one year which is the forward-looking time horizon used here.

In the time difference regression in Figure 1, the efficiency residuals are weighted zonewise by their market share, i.e. individual output divided by zonal output. Another approach is to use Domar weights, i.e. individual output divided by zonal value-added. Domar weights should normally be used when aggregating productivity to higher levels of the economy in order to account for the double effect of efficiency improvements—both on the own operations of the firm and on the operations on other firms using its products as intermediate inputs (Schreyer, 2001; Hulten, 1978; Domar, 1961). However, in a spatial setting it is not obvious if this should be done. It seems that one prerequisite for doing this should be that both the producer and the intermediate input user are located together in the same zone. If not, the productivity gain of the output of the producing firm should accrue to the intermediate input user in another zone, where that firm is located (if that by any chance would be known to the researcher). In any case, this is a potential source of underestimation of the aggregated productivity gains. To my knowledge this issue has never been investigated, and nowhere in the literature on firm productivity have Domar weights been used so far.[18]

# 7 Acknowledgements

I would like to thank Hans Lööf for comments on an earlier draft, and the Swedish Governmental Agency for Innovation Systems (VINNOVA) and Centre for Transport Studies at KTH for financial support.

# References

Åkerman, A. (2009, May). Trade, reallocations and productivity: A bridge between theory and data in Öresund. Draft paper, Dept. of Economics, Stockholm University. Url: http://people.su.se/ ank/Akerman_EIIE.pdf. 32

Ackerberg, D., C. L. Benkard, S. Berry, and A. Pakes (2007). Econometric tools for analyzing market outcomes. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6A, Chapter 63, pp. 4171–4276. Amsterdam: North-Holland. 14, 15, 25

Ackerberg, D. A., K. Caves, and G. Frazer (2006, December). Structural identification of production functions. Mimeo, Dept. of Economics, University of California, Los Angeles. Url: http://www.econ.ucla.edu/ackerber/. 26

Arellano, M. and S. R. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies 58*, 277–297. 3

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-component models. *Journal of Econometrics 68*, 29–51. 3

Aschauer, D. A. (1989). Is public expenditure productive? *Journal of Monetary Economics 23*, 177–200. 1

[18]However, see Petrin and Levinsohn (2010) for an extension of the traditional aggregation procedures of firm productivity.

Bauer, P. W. (1990). Decomposing TFP growth in the presence of cost inefficiency, nonconstant returns to scale, and technological progress. *The Journal of Productivity Analysis 1*, 287–299. 5

Bell, M. E. and T. J. McGuire (Eds.) (1997). *Macroeconomic analysis of the linkages between transportation investments and economic performance*, NCHRP (National Cooperative Highway Research Program), Washington, D.C. Transportation Research Board, National Research Council: National Academy Press. Report 389. 2

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*, 115–143. 3, 31

Bom, P. and J. E. Ligthart (2009, August). How productive is public capital? A meta-analysis. Working paper 09-12. 2

Canning, D. and E. Bennathan (2000, July). The social rate of return on infrastructure investments. Policy Research Working Paper 2390, World Bank. 32

Canning, D. and M. Fay (1993, May). The effect of transportation networks on economic growth. Economics Discussion Paper 9293-653a. 32

Domar, E. D. (1961, December). On the measurement of technological change. *The Economic Journal 71*(284), pp. 709–729. 32

Eberhardt, M. and C. Helmers (2010, November). Untested assumptions and data slicing: A critical review of firm-level production function estimators. Oxford Economic Papers WP 513. 3

Ericson, R. and A. Pakes (1995). Markov-perfect industry dynamics: A framework for empirical work. *Review of Economic Studies 62*, 53–82. 16

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A 120*(3), 253–290. 5

Gandhi, A., S. Navarro, and D. Rivers (2009, December). Identifying production functions using restrictions from economic theory. draft paper. 4

Gill, I. and C.-C. Goh (2010). Scale economies and cities. *World Bank Research Observer 25*(2), 235–262. 2

Gorodnichenko, Y. (2007, December). Using firm optimization to evaluate and estimate returns to scale. Working paper 13666, National Bureau of Economic Research (NBER), Cambridge, MA. 6

Gramlich, E. M. (1994, September). Infrastructure investment: A review essay. *Journal of Economic Literature 32*, 1176–1196. 2

Greene, W. H. (2000). *Econometric Analysis* (4th ed.). New Jersey: Prentice Hall. 4

Griliches, Z. (1957). Specification bias in estimates of production functions. *Journal of Farm Economics 39*, 8–20. 22

Griliches, Z. and J. A. Hausman (1986). Errors in variables in panel data. *Journal of Econometrics 31*, 93–118. 3

Griliches, Z. and J. Mairesse (1995, March). Production functions: The search for identification. NBER Working Papers 5067, National Bureau of Economic Research, Inc. 3

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*, 1029–1054. 3

Hoch, I. (1958). Simultaneous equation bias in the context of the Cobb-Douglas production function. *Econometrica 26*(4), 566–578. 6

Hsu, C.-H., J. M. G. Taylor, S. Murray, and D. Commenges (2007). Multiple imputation for interval censored data with auxiliary variables. *Statistics in Medicine 26*(4), 769–781. 14

Hulten, C. R. (1978). Growth accounting with intermediate inputs. *The Review of Economic Studies 45*(3), pp. 511–518. 32

Lakshmanan, T. (2010). The broader economic consequences of transport infrastructure investments. *Journal of Transport Geography In Press, Corrected Proof.* 2

Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies 70*, 317–341. 14, 16

Mamuneas, T. P. and M. I. Nadiri (2006, August). Production, consumption and the rates of return to highway infrastructure capital. Paper presented at the 62nd Annual Congress of the International Institute of Public Finance (IIPF). 32

Marschak, J. and W. H. Andrews, Jr (1944, July-Oct). Random simultaneous equations and the theory of production. *Econometrica 12*(3–4), 143–205. Cowles Commission Papers, New Series, No. 5. 3

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica, 71*(6), 1695–1725. 32

Melo, P. C., D. J. Graham, and R. B. Noland (2009). A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics 39*(3), 332 – 342. 2

Mera, K. (1973). On the urban agglomeration and economic efficiency. *Economic Development and Cultural Change 21*(2), 309–324. 2

Mikelbank, B. A. and R. W. Jackson (2000, July). The role of space in public capital research. *International Regional Science Review 23*(3), 235–258. 2

Mishra, S. K. (2007, October). A brief history of production functions. Working paper, Dept. of Economics, Shillong, India. Url: http://ssrn.com/paper=1020577. 17

Muendler, M.-A. (2007, October). Estimating production functions when productivity change is endogenous. Mimeo, Univ. of California, San Diego. Url: http://econ.ucsd.edu/muendler/papers/tfpest.pdf. 14

Munnell, A. H. (Ed.) (1990, June). *Is there a shortfall in public capital investment?* Federal Reserve Bank of Boston. Conference Series No. 34, http://www.bos.frb.org/economic/conf/conf34/index.htm. 1

Munnell, A. H. (1992). Policy watch: Infrastructure investment and economic growth. *Journal of Economic Perspectives 6*(4), 189–198. 1

Olley, G. S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica 64*(6), 1263–1297. 1, 5, 14, 21, 22, 25, 29, 31

Pakes, A. (1994). The estimation of dynamic structural models:. In J. Laffont and C. Sims (Eds.), *Proceedings of the 6th World Congress of the Econometric Society*, Advances in Econometrics, Chapter 5, pp. 171–259. Econometric Society. 14, 17

Pakes, A. and Z. Griliches (1984). Estimating distributed lags in short panels with an application to the specification of depreciation patterns and capital stock constructs. *The Review of Economic Studies 51*(2), 243–262. 11

Pan, W. (2001). A multiple imputation approach to regression analysis for doubly censored data with application to aids studies. *Biometrics 57*(4), 1245–1250. 14

Petersen, T. (2004a, May). Estimating the link between accessibility and productivity with propensity score matching. In *Modelling transport, accessibility and productivity in Öresund*, Licentiate thesis, pp. 133–168. Stockholm: Dept. of Infrastructure, KTH. TRITA-INFRA 04-014. 31

Petersen, T. (2004b, May). The impact of accessibility on cost: A microdata panel approach. In *Modelling transport, accessibility and productivity in Öresund*, Licentiate thesis, pp. 23–129. Stockholm: Department of Infrastructure, KTH. TRITA-INFRA 04-014. 31

Petrin, A. and J. Levinsohn (2010, May). Measuring aggregate productivity growth using plant-level data. Unpublished manuscript, University of Minnesota, Twin Cities. Url: http://www.econ.umn.edu/~petrin/papers/pl23_ap.pdf. 33

Romp, W. and J. de Haan (2007). Public capital and economic growth: A critical survey. *Perspektiven der Wirtschaftspolitik 8*(S1), 6–52. 2

Schreyer, P. (2001). Measuring productivity: Measurement of aggregate and industry-level productivity growth. Manual, Organisation for Economic Cooperation and Development, Paris, France. 32

Skytesvall, T. and H.-O. Hagen (2006, 16–18 October). Does Sweden give away its TFP for free? In E. Diewert (Ed.), *OECD Workshop on Productivity Analysis and Measurement*. Bern: OECD Statistics Directorate. 32

Straub, S. (2008, January). Infrastructure and growth in developing countries: Recent advances and research challenges. Policy Research Working Paper 4460, The World Bank. 2

Sturm, J.-E., G. H. Kuper, and J. de Haan (1998). Modelling government investment and economic growth on a macro level: A review. In S. Brakman, H. van Ees, and S. Kuipers (Eds.), *Market behaviour and macroeconomic modelling*, Chapter 14, pp. 359–406. Basingstoke, Hampshire: Palgrave Macmillan. 2

Syverson, C. (2010, April). What determines productivity? Unpublished manuscript, Booth School of Business, University of Chicago, Chicago. http://home.uchicago.edu/~syverson/productivitysurvey.pdf. 4

Tatom, J. A. (1991, May/June). Public capital and private sector performance. *Federal Reserve Bank of St. Louis Review 73*(3), 3–15. 1

Tatom, J. A. (1993, August). Paved with good intentions: The mythical national infrastructure crisis. *Cato Policy Analysis* (196). Accessed December 21, 2010. 1

# A Appendix

## A.1 Data summary

| SNI | Obs | Acc/Inv missing (%) | Firms | Avg prod (mSEK) | Avg empl. | inv $\leqslant 0$ (%) | Avg net inv (kSEK) | Avg grs inv (mSEK) |
|---|---|---|---|---|---|---|---|---|
| 1 | 6,506 | 53 | 1,135 | 5.1 | 4.2 | 18 | 427.8 | 185.2 |
| 15 | 1,682 | 32 | 348 | 33.1 | 18.6 | 16 | 1,419.6 | 173.9 |
| 20 | 1,239 | 48 | 247 | 17.1 | 11.6 | 23 | 593.1 | 131.9 |
| 22 | 3,242 | 35 | 651 | 13.1 | 9.5 | 22 | 476.8 | 157.4 |
| 24 | 856 | 35 | 156 | 113.5 | 40.8 | 15 | 4,689.1 | 544.3 |
| 25 | 991 | 34 | 192 | 45.3 | 25.5 | 15 | 2,089.8 | 480.1 |
| 28 | 4,265 | 35 | 776 | 10.4 | 9.3 | 18 | 432.6 | 88.7 |
| 29 | 2,622 | 38 | 492 | 19.6 | 13.5 | 18 | 564.3 | 101.5 |
| 31 | 697 | 36 | 121 | 20.1 | 16.0 | 19 | 531.6 | 95.0 |
| 33 | 1,479 | 33 | 268 | 31.2 | 14.7 | 23 | 774.0 | 505.2 |
| 36 | 1,121 | 40 | 247 | 16.3 | 11.8 | 19 | 464.3 | 83.2 |
| 40 | 214 | 30 | 55 | 239.5 | 37.2 | 11 | 16,863.4 | 4,824.3 |
| 45 | 16,696 | 36 | 3,173 | 6.2 | 5.9 | 25 | 222.1 | 95.8 |
| 50 | 6,893 | 37 | 1,290 | 17.9 | 5.0 | 22 | 188.7 | 88.8 |
| 51 | 22,126 | 39 | 4,654 | 25.5 | 6.4 | 25 | 281.4 | 95.3 |
| 52 | 20,782 | 40 | 4,589 | 10.0 | 5.2 | 31 | 126.6 | 39.5 |
| 55 | 7,532 | 38 | 1,908 | 4.8 | 6.5 | 23 | 205.7 | 74.9 |
| 60 | 8,421 | 40 | 1,572 | 5.9 | 5.7 | 25 | 546.6 | 186.2 |
| 63 | 2,229 | 37 | 523 | 28.1 | 8.1 | 23 | 418.6 | 113.2 |
| 70 | 4,382 | 42 | 1,214 | 7.9 | 3.7 | 28 | 2,185.5 | 1,752.6 |
| 71 | 1,285 | 44 | 340 | 8.3 | 3.9 | 25 | 982.8 | 2,187.2 |
| 72 | 4,613 | 37 | 1,270 | 8.6 | 7.4 | 23 | 226.1 | 55.3 |
| 73 | 904 | 38 | 231 | 85.5 | 14.6 | 19 | 1,227.7 | 209.2 |
| 74 | 30,166 | 38 | 6,794 | 4.9 | 4.4 | 26 | 147.0 | 92.3 |
| 80 | 1,937 | 36 | 582 | 3.6 | 5.9 | 24 | 165.4 | 45.4 |
| 85 | 7,698 | 34 | 1,407 | 3.9 | 7.0 | 21 | 154.7 | 42.5 |
| 92 | 2,843 | 44 | 723 | 4.5 | 5.4 | 23 | 254.4 | 92.1 |
| 93 | 1,719 | 37 | 375 | 2.3 | 3.7 | 32 | 94.6 | 26.6 |
| Goods | 45,399 | 39 | 8,369 | 16.1 | 10.1 | 21 | 689.0 | 180.3 |
| Service | 126,818 | 39 | 27,099 | 12.0 | 5.8 | 26 | 331.4 | 192.0 |
| All | 174,484 | 39 | 35,295 | 13.6 | 7.2 | 24 | 441.7 | 192.1 |

Table 9: Summary statistics per industry. A key to the industries is found in Table 20 in the Appendix.

## A.2 Firm dynamics

| SNI | surv | median age | start | chg loc | chg act | chg own | foreign |
|---|---|---|---|---|---|---|---|
| 1 | 0.95 | 9.1 | 0.015 | 0.030 | 0.089 | 0.004 | 0.005 |
| 15 | 0.94 | 9.4 | 0.037 | 0.032 | 0.046 | 0.015 | 0.026 |
| 20 | 0.94 | 9.2 | 0.031 | 0.036 | 0.025 | 0.007 | 0.012 |
| 22 | 0.92 | 9.9 | 0.024 | 0.047 | 0.023 | 0.022 | 0.036 |
| 24 | 0.96 | 9.2 | 0.026 | 0.026 | 0.036 | 0.022 | 0.164 |
| 25 | 0.96 | 9.9 | 0.025 | 0.025 | 0.057 | 0.031 | 0.064 |
| 28 | 0.95 | 9.8 | 0.026 | 0.037 | 0.030 | 0.012 | 0.013 |
| 29 | 0.95 | 10.3 | 0.020 | 0.043 | 0.051 | 0.026 | 0.038 |
| 31 | 0.96 | 9.0 | 0.014 | 0.024 | 0.055 | 0.023 | 0.039 |
| 33 | 0.95 | 9.4 | 0.023 | 0.032 | 0.026 | 0.015 | 0.034 |
| 36 | 0.93 | 8.4 | 0.046 | 0.038 | 0.042 | 0.029 | 0.024 |
| 40 | 0.90 | 6.3 | 0.037 | 0.009 | 0.103 | 0.075 | 0.070 |
| 45 | 0.94 | 8.9 | 0.033 | 0.043 | 0.014 | 0.009 | 0.005 |
| 50 | 0.94 | 8.9 | 0.031 | 0.037 | 0.012 | 0.009 | 0.015 |
| 51 | 0.92 | 8.8 | 0.030 | 0.049 | 0.049 | 0.021 | 0.058 |
| 52 | 0.92 | 8.6 | 0.037 | 0.034 | 0.015 | 0.013 | 0.008 |
| 55 | 0.90 | 6.4 | 0.059 | 0.032 | 0.028 | 0.012 | 0.009 |
| 60 | 0.94 | 8.4 | 0.032 | 0.048 | 0.007 | 0.005 | 0.004 |
| 63 | 0.92 | 6.4 | 0.056 | 0.061 | 0.020 | 0.029 | 0.051 |
| 70 | 0.91 | 7.3 | 0.040 | 0.044 | 0.047 | 0.024 | 0.022 |
| 71 | 0.89 | 7.9 | 0.038 | 0.043 | 0.030 | 0.015 | 0.024 |
| 72 | 0.89 | 5.4 | 0.071 | 0.063 | 0.104 | 0.026 | 0.022 |
| 73 | 0.93 | 5.8 | 0.054 | 0.034 | 0.033 | 0.028 | 0.065 |
| 74 | 0.92 | 7.4 | 0.038 | 0.049 | 0.024 | 0.014 | 0.016 |
| 80 | 0.92 | 6.3 | 0.059 | 0.054 | 0.118 | 0.020 | 0.007 |
| 85 | 0.96 | 7.6 | 0.029 | 0.036 | 0.068 | 0.008 | 0.004 |
| 92 | 0.90 | 7.0 | 0.044 | 0.042 | 0.022 | 0.013 | 0.011 |
| 93 | 0.91 | 8.3 | 0.040 | 0.029 | 0.008 | 0.006 | 0.002 |
| Goods | 0.94 | 9.2 | 0.027 | 0.038 | 0.036 | 0.013 | 0.021 |
| Services | 0.92 | 7.9 | 0.038 | 0.044 | 0.034 | 0.015 | 0.023 |
| All | 0.93 | 8.4 | 0.034 | 0.042 | 0.034 | 0.014 | 0.023 |

Table 10: Average dynamics over all firms and time periods per industry. Average rates (survival from one year to the next, new (starting) firms, change of location, change of activity, change of ownership), median age, and frequency of foreign ownership.

## A.3 Results

### A.3.1 Survival model—elasticities

| | NetInvVol | BldCap | MachCap | Acc1669Tot | VAPEmpl | VAPTO | Labour | EPEmpl | IntDebts | Solidity | d_foreign |
|---|---|---|---|---|---|---|---|---|---|---|---|
| SNI 1 | | | 0.022 | -0.020 | 0.042 | 0.027 | 0.023 | -0.025 | | | |
| SNI 15 | | | | | | | | | | | |
| SNI 20 | | | | | | | | | | | |
| SNI 22 | | | | | 0.044 | | | | | 0.009 | |
| SNI 24 | | | | | | | | | | 0.019 | |
| SNI 25 | 0.059 | -0.029 | | | 0.110 | | 0.039 | -0.029 | | | -0.005 |
| SNI 28 | | | | | 0.049 | | | | | 0.008 | |
| SNI 29 | | | | | | | | | | | |
| SNI 31 | | | | | | | | | -0.021 | | |
| SNI 33 | | | | 0.077 | | | | | | | |
| SNI 36 | | | | | | | | | | | |
| SNI 40 | | | | | | | | | | | |
| SNI 45 | | | | -0.031 | 0.050 | 0.028 | 0.011 | -0.013 | | 0.013 | -0.000 |
| SNI 50 | | | | -0.023 | 0.034 | 0.026 | 0.021 | -0.012 | | 0.010 | |
| SNI 51 | | | -0.002 | -0.018 | 0.087 | 0.001 | 0.006 | -0.030 | | 0.006 | |
| SNI 52 | | | -0.009 | -0.027 | 0.061 | | 0.038 | -0.010 | | 0.006 | |
| SNI 55 | | | | | 0.054 | 0.060 | 0.037 | -0.015 | | | |
| SNI 60 | | | | -0.028 | 0.024 | | 0.013 | -0.009 | -0.007 | | |
| SNI 63 | | | | | 0.068 | 0.039 | 0.038 | -0.022 | | 0.018 | |
| SNI 70 | | | | -0.038 | 0.048 | 0.014 | | -0.039 | | | |
| SNI 71 | | | | | | | | | | 0.011 | |
| SNI 72 | | | | | 0.069 | | | | -0.004 | | -0.003 |
| SNI 73 | | | | | 0.050 | | | 0.012 | | | |
| SNI 74 | | 0.001 | | -0.021 | 0.069 | | | -0.014 | | 0.004 | -0.001 |
| SNI 80 | | | 0.035 | | 0.069 | 0.038 | | -0.022 | | | |
| SNI 85 | | | | -0.020 | 0.056 | | | -0.013 | -0.002 | | -0.000 |
| SNI 92 | | | | | 0.115 | | | -0.036 | | | |
| SNI 93 | | | | | 0.097 | | | -0.034 | | | -0.001 |

Table 11: Survival model: significant elasticities per industry.

### A.3.2 Survival model—marginal effects

| | age | d_chgloc | d_chgown | d_start |
|---|---|---|---|---|
| SNI 1 | | | | -8.3 |
| SNI 15 | 0.27 | | | |
| SNI 20 | | | | |
| SNI 22 | 0.32 | | | |
| SNI 24 | | | | |
| SNI 25 | | | | |
| SNI 28 | 0.30 | -4.4 | | -4.3 |
| SNI 29 | | -4.0 | | |
| SNI 31 | | | | |
| SNI 33 | | | | |
| SNI 36 | 0.84 | | | |
| SNI 40 | | | | |
| SNI 45 | 0.33 | | | |
| SNI 50 | 0.18 | | | -4.0 |
| SNI 51 | 0.33 | -2.9 | | -2.8 |
| SNI 52 | 0.35 | -7.3 | | -3.2 |
| SNI 55 | 1.02 | -6.8 | | -3.8 |
| SNI 60 | 0.30 | | | |
| SNI 63 | 0.60 | -7.5 | | |
| SNI 70 | 0.36 | | | |
| SNI 71 | | | | |
| SNI 72 | 0.84 | | | |
| SNI 73 | 0.94 | | | |
| SNI 74 | 0.35 | -2.2 | | |
| SNI 80 | 0.48 | | | |
| SNI 85 | 0.15 | | | |
| SNI 92 | 0.57 | | | |
| SNI 93 | 0.55 | | | |

Table 12: Survival model: significant marginal effects per industry (%).

### A.3.3 Production functions—Goods sector

| | Linear in $\hat{h}$ and $\hat{P}$ | | Non-linear with only $\hat{h}$ | | Non-linear with only $\hat{P}$ | | Non-linear with both $\hat{h}$ and $\hat{P}$ | |
|---|---|---|---|---|---|---|---|---|
| | Est. | Std. Err. | Est. | Std. Err. | Est. | Std. Err. | Est. | Std. Err. |
| $lm$ | 0.6738 | 0.0021 | | | | | | |
| $ll$ | 0.2935 | 0.0027 | | | | | | |
| $lkb$ | 0.0011 | 0.0004 | 0.0016 | 0.0004 | 0.0027 | 0.0004 | 0.0015 | 0.0004 |
| $lkm$ | 0.0155 | 0.0014 | 0.0229 | 0.0012 | 0.0333 | 0.0011 | 0.0208 | 0.0013 |
| $age$ | -0.0012 | 0.0003 | -0.0012 | 0.0003 | -0.0011 | 0.0004 | -0.0032 | 0.0004 |
| $lacc$ | 0.0161 | 0.0038 | 0.0208 | 0.0037 | 0.0309 | 0.0039 | 0.0298 | 0.0038 |
| $tt$ | 0.0042 | 0.0033 | 0.0022 | 0.0033 | 0.0164 | 0.0033 | 0.0060 | 0.0034 |
| $tt^2$ | -0.0003 | 0.0003 | -0.0002 | 0.0003 | -0.0010 | 0.0003 | -0.0005 | 0.0003 |
| | **Tests** | | | | | | | |
| | statistic | p-value | statistic | p-value | statistic | p-value | statistic | p-value |
| DW | 1.9154 | 0.0000 | 2.1036 | 1.0000 | 2.1036 | 1.0000 | 2.1138 | 1.0000 |
| Woold. | 27.1387 | 0.0000 | 9.2030 | 0.0000 | 9.5466 | 0.0000 | 9.0992 | 0.0000 |

Table 13: Goods - Estimates and tests for serial correlation and individual effects.

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| $lm$ | 0.6738 | 0.0021 | 326.66 | 0.000 |
| $ll$ | 0.2935 | 0.0027 | 108.80 | 0.000 |
| $lkb$ | 0.0018 | 0.0004 | 4.11 | 0.000 |
| $lkm$ | 0.0238 | 0.0014 | 16.85 | 0.000 |
| $age$ | -0.0036 | 0.0004 | -9.19 | 0.000 |
| $lacc$ | 0.0353 | 0.0040 | 8.79 | 0.000 |
| $tt$ | 0.0057 | 0.0034 | 1.71 | 0.088 |
| $tt^2$ | -0.0005 | 0.0003 | -1.53 | 0.127 |
| $\gamma_{lm}$ | -0.0134 | 0.0022 | -6.05 | 0.000 |
| $\gamma_{ll}$ | 0.0052 | 0.0030 | 1.76 | 0.078 |

Table 14: Goods - Test for significance of lagged values of $lm$ and $ll$ ($\gamma_{lm}$ and $\gamma_{ll}$). Dependent variable: $ly - \widehat{\beta}_m lm - \widehat{\beta}_l ll$.

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| $lm$ | 0.6738 | 0.0021 | 326.66 | 0.000 |
| $ll$ | 0.2935 | 0.0027 | 108.80 | 0.000 |
| $lkb$ | 0.0006 | 0.0007 | 0.86 | 0.388 |
| $lkm$ | 0.0207 | 0.0020 | 10.16 | 0.000 |
| $age$ | -0.0022 | 0.0026 | -0.85 | 0.394 |
| $lacc$ | 0.0548 | 0.0267 | 2.05 | 0.040 |
| $tt$ | 0.0050 | 0.0035 | 1.43 | 0.152 |
| $tt^2$ | -0.0004 | 0.0003 | -1.24 | 0.216 |
| $\gamma_{lkb}$ | 0.0011 | 0.0007 | 1.63 | 0.104 |
| $\gamma_{lkm}$ | 0.0001 | 0.0021 | 0.04 | 0.971 |
| $\gamma_{age}$ | -0.0010 | 0.0027 | -0.36 | 0.720 |
| $\gamma_{lacc}$ | -0.0254 | 0.0273 | -0.93 | 0.351 |

Table 15: Goods - Test for significance of lagged values of $lkb$, $lkm$, $age$ and $lacc$ ($\gamma_{lkb}$ etc.). Dependent variable: $ly - \widehat{\beta}_m lm - \widehat{\beta}_l ll$.

### A.3.4 Production functions—Service sector

| | Linear in $\hat{h}$ and $\hat{P}$ | | Non-linear with only $\hat{h}$ | | Non-linear with only $\hat{P}$ | | Non-linear with both $\hat{h}$ and $\hat{P}$ | |
|---|---|---|---|---|---|---|---|---|
| | Est. | Std. Err. | Est. | Std. Err. | Est. | Std. Err. | Est. | Std. Err. |
| $lm$ | 0.6334 | 0.0012 | | | | | | |
| $ll$ | 0.3261 | 0.0021 | | | | | | |
| $lkb$ | 0.0030 | 0.0004 | 0.0031 | 0.0004 | 0.0082 | 0.0004 | 0.0035 | 0.0004 |
| $lkm$ | 0.0062 | 0.0011 | 0.0126 | 0.0011 | 0.0148 | 0.0010 | 0.0043 | 0.0011 |
| $age$ | -0.0097 | 0.0003 | -0.0089 | 0.0004 | -0.0170 | 0.0003 | -0.0209 | 0.0004 |
| $lacc$ | 0.0473 | 0.0035 | 0.0425 | 0.0035 | 0.0909 | 0.0035 | 0.0923 | 0.0034 |
| $tt$ | -0.0064 | 0.0028 | -0.0020 | 0.0029 | 0.0043 | 0.0028 | 0.0003 | 0.0027 |
| $tt^2$ | 0.0009 | 0.0003 | 0.0002 | 0.0003 | 0.0000 | 0.0003 | 0.0002 | 0.0003 |
| | Tests | | | | | | | |
| | statistic | p-value | statistic | p-value | statistic | p-value | statistic | p-value |
| DW | 1.9605 | 0.0000 | 2.1251 | 1.0000 | 2.1251 | 1.0000 | 2.2324 | 1.0000 |
| Woold. | 42.1251 | 0.0000 | 21.3168 | 0.0000 | 15.5201 | 0.0000 | 15.4721 | 0.0000 |

Table 16: Service - Estimates and tests for serial correlation and individual effects.

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| $lm$ | 0.6334 | 0.0012 | 538.51 | 0.000 |
| $ll$ | 0.3261 | 0.0021 | 153.18 | 0.000 |
| $lkb$ | 0.0035 | 0.0004 | 8.69 | 0.000 |
| $lkm$ | 0.0092 | 0.0012 | 7.53 | 0.000 |
| $age$ | -0.0214 | 0.0004 | -51.90 | 0.000 |
| $lacc$ | 0.0970 | 0.0035 | 27.94 | 0.000 |
| $tt$ | -0.0002 | 0.0027 | -0.06 | 0.950 |
| $tt^2$ | 0.0002 | 0.0003 | 0.77 | 0.442 |
| $\gamma_{lm}$ | -0.0027 | 0.0012 | -2.24 | 0.025 |
| $\gamma_{ll}$ | -0.0157 | 0.0022 | -7.10 | 0.000 |

Table 17: Service - Test for significance of lagged values of $lm$ and $ll$ ($\gamma_{lm}$ and $\gamma_{ll}$). Dependent variable: $ly - \widehat{\beta}_m lm - \widehat{\beta}_l ll$.

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| $lm$ | 0.6334 | 0.0012 | 538.51 | 0.000 |
| $ll$ | 0.3261 | 0.0021 | 153.18 | 0.000 |
| $lkb$ | 0.0014 | 0.0006 | 2.49 | 0.013 |
| $lkm$ | 0.0152 | 0.0016 | 9.56 | 0.000 |
| $age$ | -0.0119 | 0.0019 | -6.19 | 0.000 |
| $lacc$ | -0.0536 | 0.0190 | -2.83 | 0.005 |
| $tt$ | 0.0039 | 0.0028 | 1.38 | 0.167 |
| $tt^2$ | -0.0002 | 0.0003 | -0.80 | 0.425 |
| $\gamma_{lkb}$ | 0.0028 | 0.0006 | 4.82 | 0.000 |
| $\gamma_{lkm}$ | -0.0160 | 0.0017 | -9.60 | 0.000 |
| $\gamma_{age}$ | -0.0104 | 0.0020 | -5.25 | 0.000 |
| $\gamma_{lacc}$ | 0.1545 | 0.0194 | 7.95 | 0.000 |

Table 18: Service - Test for significance of lagged values of $lkb$, $lkm$, $age$ and $lacc$ ($\gamma_{lkb}$ etc.). Dependent variable: $ly - \widehat{\beta}_m lm - \widehat{\beta}_l ll$.

### A.3.5 Production functions—All firms

| | Linear in $\hat{h}$ and $\hat{P}$ | | Non-linear with only $\hat{h}$ | | Non-linear with only $\hat{P}$ | | Non-linear with both $\hat{h}$ and $\hat{P}$ | |
|---|---|---|---|---|---|---|---|---|
| | Estimate | Std. Error | Estimate | Std. Error | Estimate | Std. Error | Estimate | Std. Error |
| $lm$ | 0.6415 | 0.0010 | | | | | | |
| $ll$ | 0.3156 | 0.0017 | | | | | | |
| $lkb$ | 0.0020 | 0.0003 | 0.0023 | 0.0003 | 0.0060 | 0.0003 | 0.0026 | 0.0003 |
| $lkm$ | 0.0059 | 0.0009 | 0.0131 | 0.0009 | 0.0158 | 0.0008 | 0.0077 | 0.0009 |
| $age$ | -0.0078 | 0.0002 | -0.0071 | 0.0003 | -0.0136 | 0.0002 | -0.0163 | 0.0003 |
| $lacc$ | 0.0486 | 0.0028 | 0.0477 | 0.0027 | 0.0972 | 0.0027 | 0.0963 | 0.0027 |
| $tt$ | -0.0023 | 0.0022 | -0.0007 | 0.0023 | 0.0103 | 0.0022 | 0.0043 | 0.0022 |
| $tt^2$ | 0.0005 | 0.0002 | 0.0001 | 0.0002 | -0.0005 | 0.0002 | -0.0002 | 0.0002 |
| Tests | | | | | | | | |
| | statistic | p-value | statistic | p-value | statistic | p-value | statistic | p-value |
| DW | 1.9453 | 0.0000 | 2.1129 | 1.0000 | 2.1129 | 1.0000 | 2.2001 | 1.0000 |
| Woold. | 48.4373 | 0.0000 | 23.9908 | 0.0000 | 20.1988 | 0.0000 | 19.7584 | 0.0000 |

Table 19: All firms - Estimates and tests for serial correlation and individual effects.

### A.4 Industry key

| | Description |
|---|---|
| SNI 1 | Agriculture, hunting and related service activities |
| SNI 15 | Manufacture of food products and beverages |
| SNI 20 | Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials |
| SNI 22 | Publishing, printing and reproduction of recorded media |
| SNI 24 | Manufacture of chemicals and chemical products |
| SNI 25 | Manufacture of rubber and plastic products |
| SNI 28 | Manufacture of fabricated metal products, except machinery and equipment |
| SNI 29 | Manufacture of machinery and equipment n.e.c. |
| SNI 31 | Manufacture of electrical machinery and apparatus n.e.c. |
| SNI 33 | Manufacture of medical, precision and optical instruments, watches and clocks |
| SNI 36 | Manufacture of furniture; manufacturing n.e.c. |
| SNI 40 | Electricity, gas, steam and hot water supply |
| SNI 45 | Construction |
| SNI 50 | Sale, maintenance and repair of motor vehicles and motorcycles; retail sale of automotive fuel |
| SNI 51 | Wholesale trade and commission trade, except of motor vehicles and motorcycles |
| SNI 52 | Retail trade, except of motor vehicles and motorcycles; repair of personal and household goods |
| SNI 55 | Hotels and restaurants |
| SNI 60 | Land transport; transport via pipelines |
| SNI 63 | Supporting and auxiliary transport activities; activities of travel agencies |
| SNI 70 | Real estate activities |
| SNI 71 | Renting of machinery and equipment without operator and of personal and household goods |
| SNI 72 | Computer and related activities |
| SNI 73 | Research and development |
| SNI 74 | Other business activities |
| SNI 80 | Education |
| SNI 85 | Health and social work |
| SNI 92 | Recreational, cultural and sporting activities |
| SNI 93 | Other service activities |

Table 20: Descriptions of included industries.