MEMORANDUM

No 30/2002

How and why do firms differ?

By
Tor Jakob Klette and Arvid Raknerud

ISSN: 0801-1117

Department of Economics
University of Oslo

This series is published by the
University of Oslo
Department of Economics

P. O.Box 1095 Blindern
N-0317 OSLO Norway
Telephone: + 47 22855127
Fax: + 47 22855035
Internet: http://www.oekonomi.uio.no/
e-mail: econdep@econ.uio.no

In co-operation with
The Frisch Centre for Economic Research

Gaustadalleén 21
N-0371 OSLO Norway
Telephone: +47 22 95 88 20
Fax: +47 22 95 88 25
Internet: http://www.frisch.uio.no/
e-mail: frisch@frisch.uio.no

List of the last 10 Memoranda:

No 29	Halvor Mehlum, Kalle Moene and Ragnar Torvik Institutions and the resource curse. 26 pp.
No 28	Jon Strand Public-good valuation and intrafamily allocation. 37 pp.
No 27	Gabriela Mundaca Optimal bailout during currency and financial crises: A sequential game analysis. 41 pp.
No 26	Halvor Mehlum At Last! An Explicit Solution for the Ramsey Saddle Path. 6 pp.
No 25	Steinar Holden and John C. Driscoll Coordination, Fair Treatment and Inflation Persistence. 37 pp.
No 24	Atle Seierstad Maximum principle for stochastic control in continuous time with hard end constraints.
No 23	Hilde C. Bjørnland and Håvard Hungnes Fundamental determinants of the long run real exchange rate: The case of Norway. 40 pp.
No 22	Atle Seierstad Conditions implying the vanishing of the Hamiltonian at the infinite horizon in optimal control problems. 3 pp.
No 21	Morten Søberg The Duhem-Quine thesis and experimental economics: A reinterpretation. 22 pp.
No 20	Erling Barth, Bernt Bratsberg and Oddbjørn Raaum Local Unemployment and the Relative Wages of Immigrants: Evidence from the Current Population Surveys. 53 pp.

A complete list of this memo-series is available in a PDF® format at:
http://www.oekonomi.uio.no/memo/

How and why do firms differ?*

Tor Jakob Klette† and Arvid Raknerud‡

ABSTRACT: How do firms differ, and why do they differ even within narrowly defined industries? Using evidence from six high-tech, manufacturing industries covering a 24-year period, we show that differences in sales, materials, labor costs and capital across firms can largely be summarized by a single, firm-specific, dynamic factor, which we label *efficiency* in the light of our structural model. The model contains the complete system of supply and factor demand equations. It suggests that efficiency is strongly linked to profitability and firm size, but it is unrelated to labor productivity. Our second task is to understand the origin and evolution of the differences in efficiency. Among the firms established within the 24 year period that we consider, permanent differences in efficiency dominate over differences generated by firm-specific, cumulated innovations.

JEL classification: C33, C51, D21
Keywords: efficiency, firm heterogeneity, labor productivity, intrinsic differences, firm-specific innovations, state space models, maximum likelihood

---

*This paper has benefited from comments at seminars at the Institute for Fiscal Studies, the University of Oslo, the University of Helsinki, the Norwegian School of Economics, the Norwegian School of Management, the Frisch Center, Statistics Norway, the University of Minnesota, an NBER Productivity Workshop, and the 2002 Econometric Society European Meeting (ESEM 2002) in Venice. Comments and suggestions by Boyan Jovanovic, Sam Kortum, Kalle Moene, Jarle Møen and Ariel Pakes are gratefully acknowledged. This research has been financially supported by The Norwegian Research Council ("Næring, Finans, Marked").

†Department of Economics, University of Oslo, and CEPR. *E-mail:* t.j.klette@econ.uio.no, `http://folk.uio.no/torjk/` .

‡Research Department, Statistics Norway. *E-mail:* arvid.raknerud@ssb.no .

1 Introduction

More than 50 years ago Marschak and Andrews (1944) showed that production function regressions generate inconsistent parameter estimates because optimal supply and factor inputs are jointly determined by unobservable differences in efficiency across firms. The problem with regressions on firm level data has haunted studies of efficiency and producer behavior ever since; see Griliches and Mairesse (1998) for a survey. In this paper, we propose an econometric model that explicitly uses the full system of equations derived from optimizing supply and factor demands to overcome this problem. The econometric model allows us to explore the origins of the efficiency differences across firms.

Efficiency differences are decomposed into stochastic, firm-specific (idiosyncratic) *cumulated innovations* as emphasized e.g. by Ericson and Pakes (1995), and *permanent efficiency differences* as emphasized by Jovanovic (1982) and others[1]. In the six high-tech industries that we examine, the efficiency differences are largely permanent. Cumulated innovations in efficiency play a lesser role among the firms established within our 24 year period.

A large literature on firm heterogeneity has focused on firm performance as measured by size (sales or employment), including Pakes and Ericson (1998). However, most recent studies of differences in firm performance have focused on differences in efficiency. In competitive environments, differences in size and efficiency should be closely related as more efficient firms will tend to be larger, see e.g. Demsetz (1973), Lucas (1978), and Jovanovic (1982). Our structural model highlights the positive relationship between size and efficiency, while also emphasizing that the fixity of capital is essential in explaining differences in firm sizes.

We use the term efficiency rather than productivity, as our structural model suggests that differences in labor productivity are unrelated to differences in efficiency. The argument is simple, but seems to have been overlooked in the literature: Consider firms with different levels of efficiency competing in a frictionless industry. A firm with high efficiency will choose a high level of labor input so that its marginal product is equal to the real wage, which, by assumption, is the same across all firms[2]. With a Cobb-Douglas

[1] Appendix A gives a survey of theoretical models focusing on firm heterogeneity.
[2] We assume diminishing returns for profit-maximization to be well defined.

production function, the marginal product is proportional to production per factor input, and, hence, all firms should have the same level of production per factor input apart from transient noise or fluctuations[3]. This argument raises the question of how to make inferences about differences in efficiency from firm level data, which is a central theme in our analysis.

Our econometric framework uses a state space-approach, in combination with the Kalman filter and smoother, to decompose the observations of firm-level supply and factor demands in terms of four types of latent components: (i) firm-specific permanent components, (ii) firm-specific stochastic trends, (iii) transient noise, and (iv) industry-wide fluctuations. The multivariate framework imposes few restrictions on the data generating process *a priori* and allows us to consider the validity of the restrictions imposed by our structural model. Our testing procedure relates to co-integrated time-series analysis. Our structural model of firm behavior implies that supply and factor inputs should be co-integrated with a heavily constrained co-integrating vector, and we show that these constraints are largely satisfied in all industries. The model is estimated by a partial likelihood function and we discuss the question of identification emphasizing sample attrition and the fact that we do not explicitly model the firms' exit decisions.

2 A first look at differences in firm performance

How should we measure differences in firm performance and do these differences increase with firm age? Using size as a preliminary measure of firm performance, we address the second question in Figure 1[4]. Figure 1 presents the means and standard deviations of log sales as a function of firm age. All observations are measured relative to industry-year means. Not surprisingly, the graph shows that on average young firms are substantially smaller than older firms and that firm growth tends to decelerate with age. More interestingly, the graph shows that relative differences in firm size are almost independent of firm age.

The upper graph in Figure 2 displays the correlation coefficient between log sales in

[3]This result also holds in the CES case and, more generally, when there is a one-to-one relation between marginal product and production per factor input.

[4]Figures 1-2 are based on a comprehensive, unbalanced sample of firm level observations from six (two-digit NACE) high-tech manufacturing industries, as discussed in Section 5. Graphs for the six separate industries show the same patterns as in Figures 1-3.

the firms' first year and in their subsequent years. The correlation coefficient for the first and the second year is 0.94, and it declines slowly in the subsequent years. This shows that the *relative* differences in firm size are highly persistent as the firms become older.

These patterns indicate that differences across young firms are as large as those among older firms and the differences are highly persistent, suggesting that firm heterogeneity is generated by permanent differences. However, this conclusion is preliminary as it leaves open a number of questions. Young firms have a high rate of exit; on average, 50 percent of a new cohort of firms have exited within seven years in our sample. Since exiting firms are systematically selected among the least successful firms, we expect an upward trend in average log sales. Such an upward trend is clearly seen in Figure 1. Systematic selection that eliminates the least successful firms should also, *cet.par.*, tend to narrow down the differences in firm size. However, such narrowing is not visible in the figure. There must be an offsetting force that tends to make firms grow more unequal with age. Such an offsetting force could be idiosyncratic, cumulated shocks that would also explain the declining correlation between a firm's performance in its first year and in its subsequent years, demonstrated in Figure 2.

Labor productivity is another widely used measure of firm performance. Figure 3 presents means and standard deviations of labor productivity as a function of firm age. We see that the patterns are rather different from those in Figure 1. There is no upward trend in labor productivity and the standard deviations decline substantially with age. The difference between sales and labor productivity is equally clear from the lower graph in Figure 2, which displays the correlation coefficient between labor productivity in the firms' first year and in their subsequent years[5]. The low correlation coefficient between productivity in the first two years shows that almost half of the observed variance in labor productivity is due to temporary fluctuations or noise in the data. A comparison of the two graphs in Figure 2 raises the question of why differences in size are considerably more persistent than differences in labor productivity. Furthermore, this comparison indicates that labor productivity is a rather noisy measure of efficiency, as we will discuss further below.

[5]Figures 1-3 focus on heterogeneity in new cohorts of firms. Similar patterns of heterogeneity and autocorrelation are also present among older and larger firms. E.g. high and low degrees of persistence in differences in revenues and labor productivity, respectively, are not restricted to the firms' early years.

3 A structural model of optimal supply and factor demand

Our preliminary look at the data suggests that we need an econometric framework that can address a number of challenging methodological issues. The framework must account for the permanent differences embedded in firms at birth and how these differences evolve over time. In addition, it must account for the considerable noise in the data and self-selection. Yet it should be flexible enough to enable us to examine alternative measures of firm performance.

Section 3.1 presents a model of optimal supply and factor demand. This model is the basis for the econometric framework that we use to make inferences about unobserved differences in efficiency from observed supply and factor demand, as explained in Section 3.2.

3.1 Optimal supply and factor demand

Consider the production function

$$Q_{it} = A_{it} K_{i,t-1}^{\gamma} F\left(M_{it}, L_{it}\right), \tag{1}$$

where Q_{it} and A_{it} denote firm i's output and efficiency in year t, $K_{i,t-1}$ is the predetermined capital stock, and $F(M_{it}, L_{it})$ is a function aggregating materials and labor inputs. $F(M_{it}, L_{it})$ is homogenous of degree ε ($\varepsilon < 1$). Given common prices across firms for output, labor and materials, $P_t = \left(p_t, w_t^l, w_t^m\right)$, it follows that the short-run cost-function has the following form:

$$C(P_t, Q_{it}, A_{it}, K_{i,t-1}) = G(P_t) \left(\frac{Q_{it}}{A_{it} K_{i,t-1}^{\gamma}} \right)^{1/\varepsilon}. \tag{2}$$

Setting price equal to marginal costs, we obtain the following set of supply and (short-run) factor demand equations:

$$\begin{bmatrix} \ln Q_{it} \\ \ln M_{it} \\ \ln L_{it} \end{bmatrix} = \begin{bmatrix} (1-\varepsilon)^{-1} \\ (1-\varepsilon)^{-1} \\ (1-\varepsilon)^{-1} \end{bmatrix} \ln A_{it} + \begin{bmatrix} \gamma (1-\varepsilon)^{-1} \\ \gamma (1-\varepsilon)^{-1} \\ \gamma (1-\varepsilon)^{-1} \end{bmatrix} \ln K_{i,t-1} + \mathbf{g}(P_t), \tag{3}$$

where $\mathbf{g}(P_t)$ is a vector function common across firms that depends (only) on the common price vector P_t. Its functional form reflects the properties of the aggregation function $F(\cdot, \cdot)$.

According to (3), differences in firm output, material use and labor input are informative about *unobserved* differences in firm efficiency, conditional on the firms' capital stocks. The equations in (3) cannot be directly exploited to make inferences about the differences in efficiency, as these tend to be (positively) correlated with differences in capital. Hence, to obtain an econometric model that allows us to make inferences about differences in efficiency, we must introduce a model of capital accumulation.

Capital stock dynamics: Consider now the capital stock dynamics derived from each firm's optimal investment behavior. Let I_{it} denote the resources required to change the firm's capital stock from $K_{i,t-1}$ at the end of period $t-1$ to K_{it} at the end of period t, while q_t denotes the price per unit I_{it}.

Assume that (A_{it}, P^t) is Markovian, where $P^t = (P_t, q_t)$. Then the firm's investment problem is the solution of the Bellman equation:

$$\begin{aligned} V(A_{it}, K_{i,t-1}, P^t) \quad = \quad & \max_{K_{it}} \{\Pi(A_{it}, K_{i,t-1}, P_t) - q_t\, I_{it} \\ & + \beta\, E[V(A_{i,t+1}, K_{it}, P^{t+1})|\Omega_{it}]\}\,, \end{aligned} \tag{4}$$

where $V(A_{it}, K_{i,t-1}, P^t)$ is the value function, and

$$\Pi(A_{it}, K_{i,t-1}, P_t) = \pi(P_t)\ (A_{it} K_{i,t-1}^{\gamma})^{1/(1-\varepsilon)} \tag{5}$$

is the short-run profit function. In equations (4) and (5), β is the discount factor, $E\left[\cdot|\Omega_{it}\right]$ is the expectation conditional on the firm's information at t, and $\pi(P_t)$ is a function of input and output prices. We assume convex adjustment costs such that

$$K_{it} = K_{i,t-1}\ \left[1 - \delta + \delta^{1-\alpha}\, (I_{it}/K_{i,t-1})^{\alpha}\right],\ \ \alpha \in (0, 1). \tag{6}$$

Small α corresponds to large adjustment costs, while $\alpha = 1$ gives the standard equation for capital accumulation without adjustment costs. Appendix C shows that with constant returns to scale, i.e. $\gamma + \varepsilon = 1$, and $K_{i,t-1} \simeq K_{it}$, an optimal capital accumulation policy satisfies:

$$\ln K_{it} = \ln K_{i,t-1} + \frac{\delta\alpha}{1-\alpha}\left[\ln v(A_{it}, P^t) + \ln(\frac{\alpha\beta}{q_t})\right], \tag{7}$$

where $v(A_{it}, P^t)$ is the expected value per unit of capital in period $t+1$, conditional on the firm's information Ω_{it}.

The function $v(A_{it}, P^t)$ is increasing in A_{it}. Moreover, as discussed in Appendix C, $v(A_{it}, P^t)$ is approximately homogenous of degree $(1-\varepsilon)^{-1}$ in A_{it}. Hence, we can approximate (7) by

$$\ln K_{it} = \kappa_k \ln K_{i,t-1} + \kappa_a \ln A_{it} + \kappa_t, \tag{8}$$

where $\kappa_a = \frac{\delta\alpha}{(1-\alpha)(1-\varepsilon)}$ and κ_t is an industry-wide time varying intercept. According to (7), $\kappa_k = 1$, but with decreasing returns to scale, the optimal investment behavior implies that $\frac{d\ln K_{it}}{d\ln K_{i,t-1}} < 1$. Thus, we have in (8) included a parameter κ_κ, which is less than one if there are decreasing returns to scale[6].

Supply and factor demand: Combining (3) and (8), we obtain a simultaneous system of equations:

$$\mathbf{y}_{it} = \boldsymbol{\theta}_a \ln A_{i1} + \boldsymbol{\theta}_a \ln (A_{it}/A_{i1}) + \boldsymbol{\theta}_k \ln (K_{i,t-1}) + \boldsymbol{\theta}_t, \tag{9}$$

where

$$\begin{aligned} \mathbf{y}_{it} &\equiv \left[\begin{array}{cccc} \ln Q_{it} & \ln M_{it} & \ln L_{it} & \ln K_{it} \end{array} \right]' \\ \boldsymbol{\theta}_a &= \left[\begin{array}{cccc} \frac{1}{1-\varepsilon}, & \frac{1}{1-\varepsilon}, & \frac{1}{1-\varepsilon}, & \kappa_a \end{array} \right]' \\ \boldsymbol{\theta}_k &= \left[\begin{array}{cccc} \frac{\gamma}{1-\varepsilon}, & \frac{\gamma}{1-\varepsilon}, & \frac{\gamma}{1-\varepsilon}, & \kappa_k \end{array} \right]' \end{aligned} \tag{10}$$

while $\boldsymbol{\theta}_t = \left[\begin{array}{cc} \mathbf{g}(P_t)', & \kappa_t \end{array} \right]'$.

The model (9)-(10) suggests that *differences* between firms in the endogenous variables $\mathbf{y}_{it}$ are due to differences in *efficiency* $\ln (A_{it})$ and *capital accumulation*, $\ln (K_{i,t-1})$. Capital accumulation, according to (7), is driven by cumulated changes in efficiency and changes in input and output prices. Equation (9) decomposes differences in efficiency into two components: permanent differences already introduced when the firms are established, $\ln A_{i1}$, and differences in subsequent innovations, i.e. the cumulated changes in efficiency, $\ln (A_{it}/A_{i1})$.

Efficiency, profitability and labor productivity: Before we complete our econometric model by specifying its stochastic properties, we discuss how our model relates differences in efficiency to profitability and labor productivity. According to (5), (short-run)

[6] However, in that case κ_k cannot be given a direct interpretation in terms of the elasticity of scale.

profitability is increasing in efficiency A_{it} and capital $K_{i,t-1}$. On the other hand, (3) shows that differences in labor productivity, i.e. value added per labor input $(Q_{it} - M_{it})/L_{it}$, are independent of differences in firm efficiency, A_{it}. This result shows that differences in efficiency and capital intensity are inadequate to explain differences in labor productivity. The relationship between various measures of size and efficiency on the one hand and the absence of a similar relationship between labor productivity and efficiency on the other, may explain why differences in sales are much more persistent than the differences in labor productivity, as we saw in Figure 2. We will elaborate on this theme in the concluding Section 9.

3.2 The econometric model

The model of firm behavior, (9)-(10), is highly constraining on the data as it assumes that efficiency changes affect all the components of $\mathbf{y}_{it}$ through a single latent variable, A_{it}, and, furthermore, that the three first components of the "loading vector" $\boldsymbol{\theta}_a$ are equal. Notice, however, that $\boldsymbol{\theta}_a$ (and consequently γ) are not identified, because A_{it} is not observed (by the econometrician).

In this section we formulate a more general econometric model that encompasses the structural model. This general econometric model imposes considerably less structure on the data generating process than (9)-(10), and allows us to test the empirical validity of the structural restrictions. Our general model is:

$$\mathbf{y}_{it} = \mathbf{v}_i + \mathbf{a}_{it} + \boldsymbol{\theta}_k \ln K_{i,t-1} + \mathbf{d}_t + \mathbf{e}_{it}, \;\; \tau_i \le t \le T, \tag{11}$$

where

$$\mathbf{a}_{it} = \begin{cases} \mathbf{0}_4 & t = \tau_i \\ \mathbf{a}_{i,t-1} + \boldsymbol{\eta}_{it} & t = \tau_i + 1, ..., T, \end{cases} \tag{12}$$

$\mathbf{0}_k$ denotes the k-dimensional vector of zeros, and $\mathbf{v}_i, \boldsymbol{\eta}_{it}$ and $\mathbf{e}_{it}$ are 4-dimensional vectors that have independent, multivariate normal distributions:

$$\mathbf{v}_i \sim \mathcal{IN}(\mathbf{0}_4, \boldsymbol{\Sigma}_v), \quad \boldsymbol{\eta}_{it} \sim \mathcal{IN}(\mathbf{0}_4, \boldsymbol{\Sigma}_\eta), \;\; \mathbf{e}_{it} \sim \mathcal{IN}(\mathbf{0}_4, \boldsymbol{\Sigma}_e). \tag{13}$$

We have an unbalanced panel data set, where firm i is observed from year $\tau_i \ge 1$ until $T_i \le T$, where τ_i is the date of the firm's birth. The birth dates τ_i have an exogenous distribution, while the exit dates T_i can be endogenous, as we discuss in Section 6.2.

When interpreting equation (11) in view of the structural equation (9), the term $\mathbf{a}_{it}$ corresponds to $\boldsymbol{\theta}_a \ln(A_{it}/A_{i1})$, $\mathbf{v}_i$ corresponds to $\boldsymbol{\theta}_a \ln(A_{i1})$, while all transient shocks and measurement errors are captured by $\mathbf{e}_{it}$. While it may seem restrictive to assume that $\mathbf{a}_{it}$ is a random walk, our econometric procedure does not critically depend on moderate departures from the random walk assumption, as discussed in Appendix B. For example, our main results (presented in Section 7) will not be seriously affected if the $\mathbf{a}_{it}$ process is slightly mean reverting, as suggested by Blundell and Bond (1999, 2000).

The structure of the covariance matrices are essential for the interpretation and identification of the model (11)-(13), which encompasses some well-known econometric models of firm heterogeneity as special cases: If $\mathbf{\Sigma}_\eta = \mathbf{0}_{4\times4}$, we obtain the fixed effect model widely used to account for firm heterogeneity in the econometric panel data literature ($\mathbf{0}_{k\times k}$ denotes the $k \times k$ matrix of zeros). When $\mathbf{\Sigma}_e = \mathbf{0}_{4\times4}$, the model is consistent with Gibrat's law discussed by Sutton (1997), where firm growth from period $t-1$ to t is independent of the level in period $t-1$. On the other hand, when $\mathbf{\Sigma}_e$ is a non-zero matrix, the model (11)-(13) implies "mean reversion", in the sense that any component of $\Delta\mathbf{y}_{it}$ will be negatively correlated with the corresponding component of $\mathbf{y}_{it-1}$[7].

Are the parameters of the covariance matrices identified? Consider a sample covering two years; $t = 1, 2$. From (11)-(13), ignoring capital for simplicity, we have:

$$\text{Cov}(\mathbf{y}_{it}, \mathbf{y}_{is}) = \begin{cases} \mathbf{\Sigma}_v + \mathbf{\Sigma}_\eta [\min(t,s) - 1] & t \neq s \\ \mathbf{\Sigma}_v + \mathbf{\Sigma}_\eta (t-1) + \mathbf{\Sigma}_e & t = s. \end{cases} \tag{14}$$

We then obtain: $\text{Cov}(\mathbf{y}_{i2}, \mathbf{y}_{i1}) = \mathbf{\Sigma}_v$, $\text{Cov}(\mathbf{y}_{i1}, \mathbf{y}_{i1}) = \mathbf{\Sigma}_v + \mathbf{\Sigma}_e$, and $\text{Cov}(\mathbf{y}_{i2}, \mathbf{y}_{i2}) = \mathbf{\Sigma}_v + \mathbf{\Sigma}_\eta + \mathbf{\Sigma}_e$. Although identification of the covariance matrices thus appears almost trivial, the situation is complicated by sample attrition, which we discuss in Section 6.2.

Testing the structural model: As mentioned, there are no a priori constraints (apart from positive semi-definiteness) on the covariance matrices $\mathbf{\Sigma}_v$ and $\mathbf{\Sigma}_\eta$ in our general econometric model (11)-(13). On the other hand, according to the structural model (9)-(10) these two matrices can be factorized as:

$$\begin{aligned} \mathbf{\Sigma}_v &= \boldsymbol{\theta}_a \boldsymbol{\theta}_a' \, Var(\ln A_{i1}) \\ \mathbf{\Sigma}_\eta &= \boldsymbol{\theta}_a \boldsymbol{\theta}_a' \, Var[\ln(A_{it}/A_{i1})]. \end{aligned} \tag{15}$$

[7] Friedman (1993) has emphasized that noise and temporary fluctuations in the data often mislead researchers to infer convergence across the units of observations when there is no convergence in the underlying, uncontaminated processes of interest. See also Quah (1993).

If (15) holds, the rank of $\mathbf{\Sigma}_\eta$ is 1, and all components of $\boldsymbol{\eta}_{it}$ are determined by a single latent factor, say η_{it}:

$$\boldsymbol{\eta}_{it} = \mathbf{u}_\eta \eta_{it}, \ \text{ with } \ \eta_{it} \sim \mathcal{IN}(0, \sigma_\eta^2), \tag{16}$$

where $\mathbf{u}_\eta$ is the eigenvector of Σ_η corresponding to the only non-zero eigenvalue σ_η^2 . The eigenvector is normalized so that $\|\mathbf{u}_\eta\| = 1$. From (12) and (16):

$$\mathbf{a}_{it} = \mathbf{u}_\eta a_{it}, \ \text{ where } \ a_{it} = \sum_{s \leq t} \eta_{is}. \tag{17}$$

Similarly, $\mathbf{v}_i$ can be expressed by a single latent factor v_i:

$$\mathbf{v}_i = \mathbf{u}_v v_i, \ \text{ with } \ v_i \sim \mathcal{IN}(0, \sigma_v^2), \tag{18}$$

where $\mathbf{u}_v$ is the (normalized) eigenvector of Σ_v, corresponding to the only non-zero eigenvalue σ_v^2.

According to (15) the (normalized) eigenvectors $\mathbf{u}_v$ and $\mathbf{u}_\eta$ should be identical:

$$\mathbf{u}_v = \mathbf{u}_\eta = \frac{\boldsymbol{\theta}_a}{\|\boldsymbol{\theta}_a\|}, \tag{19}$$

which is a testable restriction. From the definition of $\boldsymbol{\theta}_a$ in (13), a further testable implication of the structural model is that the first three components within each eigenvector are equal.

Preceding a test of the structure of $\mathbf{u}_\eta$ and $\mathbf{u}_v$, we must examine a more basic question: How well does a model with only one latent component - i.e. where the rank of $\mathbf{\Sigma}_v$ and $\mathbf{\Sigma}_\eta$ is one - fit the data compared with a model with no structural constraints on $\mathbf{\Sigma}_v$ and $\mathbf{\Sigma}_\eta$? Consider a $\mathbf{\Sigma}_\eta$-matrix with rank $r \leq 4$. The innovations $\boldsymbol{\eta}_{it}$ can then be represented through an orthogonal factor decomposition (see Anderson, 1984):

$$\boldsymbol{\eta}_{it} = \mathbf{u}_{\eta,(1)} \eta_{it,(1)} + \ldots + \mathbf{u}_{\eta,(r)} \eta_{it,(r)}, \tag{20}$$

where $\mathbf{u}_{\eta,(j)}$ is the normalized eigenvector of $\mathbf{\Sigma}_\eta$ corresponding to its j'th largest eigenvalue $\sigma^2_{\eta,(j)}$. Furthermore, $\eta_{it,(j)} \sim \mathcal{IN}(0, \sigma^2_{\eta,(j)})$. According to our structural model, $r = 1$, so that only the first eigenvalue is non-zero. That is, $\sigma^2_{\eta,(1)} > 0$ and $\sigma^2_{\eta,(j)} = 0$ for $j \geq 2$. Hence, if our structural model is valid, the largest eigenvalue $\widehat{\sigma}^2_{\eta,(1)}$ of the *estimated* covariance matrix $\widehat{\mathbf{\Sigma}}_\eta$ should be large relative to the others. A similar result should hold with regard to the magnitude of the estimated eigenvalues $\widehat{\sigma}^2_{v,(j)}$ of Σ_v.

Our testing procedure can be related to time series analysis and terminology. Our structural model imposes a cointegration relationship between the components of $\mathbf{y}_{it}$, with an *a priori* highly constrained cointegration vector: a linear combination $\boldsymbol{\lambda}'\mathbf{y}_{it}$ will be a stationary variable (relative to the industry-wide trend $\mathbf{d}_t$) if $\boldsymbol{\lambda}'\boldsymbol{\theta}_a = 0$.

4 Why do firms differ in efficiency?

Given the validity of our structural model, we can address questions of why firms differ. In particular, our econometric framework allows us to decompose differences in efficiency and to *quantify* the relative importance of permanent differences and cumulated innovations. A natural measure of the importance of permanent differences relative to idiosyncratic innovations in a particular year, say T, is

$$V \equiv \frac{Var\{\ln A_{i1}\}}{Var\{\ln (A_{iT}/A_{i1})\}}.$$

Note that V is identified even if $\ln A_{it}$ is not: From (17) and (18) it follows that

$$V = \frac{Var\{v_i\}}{Var\{a_{iT}\}} = \frac{\sigma_v^2}{\bar{T}\ \sigma_\eta^2}, \tag{21}$$

where σ_v^2 and σ_η^2 are the (non-zero) eigenvalues of Σ_v and Σ_η, respectively, and $\bar{T} \equiv E\{T - \tau_i\}$, i.e. the average life-time of firms operating in year T.

The measure V, defined in (21), ignores endogenous exit, which will tend to reduce the variances both in v_i and a_{iT} among the firms operating in year T. Hence, we focus on a modified version of (21): Let M_T be the set of firms that operate in year T. We define the *conditional variance* ratio, CV, as

$$CV = \frac{Var\{v_i | i \in M_T\}}{Var\{a_{iT} |\ i \in M_T\}}. \tag{22}$$

As we shall see in Section 6, CV is computed from the distribution of the latent components v_i and a_{iT} *conditional* on the observations $(\mathbf{y}_{i,\tau_i}, ..., \mathbf{y}_{i,T})$. Thus, while V is computed from the *unconditional* distribution of the latent variables, CV is calculated from their conditional distribution given the observed data. This implies that CV is considerably less sensitive to the *a priori* assumption of a random walk process for a_{it}, as it is essentially a semi-parametric measure. We will return to this issue in Section 6.3, where we also elaborate upon our discussion of the self-selection problem and other econometric issues.

5 Data and variable construction

We rely on raw data from Statistics Norway's Annual Manufacturing Census, which provide annual observations on sales, intermediates, wage costs, gross investment and other variables for all Norwegian manufacturing establishments for the period 1973-1996. The Census is comprehensive in the sense that a firm is included as soon as it starts to pay payroll taxes. Separate estimates are presented for six different industry groups corresponding to the 2-digit NACE codes; see Appendix D.

Following Caves' (1998) survey of empirical findings on firm growth and turnover, we have not stressed the distinction between a firm and an establishment[8]. The unit of observation in our data is an establishment in a given year. For convenience, we have labeled the unit a firm rather than an establishment, which is not misleading in a large majority of cases, since only 10-20 percent of the establishments belong to multi-establishment firms in the sectors we consider[9].

All costs and revenues are measured in nominal prices, and incorporate taxes and subsidies. We have not deflated the variables with the available industry wide deflators as the econometric model contains an industry wide time varying intercept vector. The model contains four variables, which are measured on log-scale: sales, labor costs, materials, and capital. Sales are adjusted for inventory changes. Labor costs incorporate salaries and wages in cash and kind, social security and other costs incurred by the employer. The capital variable is constructed on the basis of annual fire insurance values and gross investment (including repairs).

Initially *all* firms in a sector that were operating during 1973-96 were included in the sample, and observed until $T = 1996$. For the firms established before 1973 we introduced separate (nuisance) parameters for the distribution of v_i[10], since v_i for these firms is composed of both permanent differences and cumulated innovations (up until 1973) and therefore has a different meaning than for firms established after 1972. For

[8] Caves (1998) points out that most of the results on firm growth and turnover have been insensitive to the establishment-firm distinction.

[9] This is not to deny that the distinction between firms (or lines-of-business) and establishments raises interesting questions for our analysis. For instance, are there strong correlations between efficiency levels across establishments within a firm? Do new establishments from an existing firm have the same efficiency as new firms? We will investigate these and related questions in future research.

[10] That is, $v_i \sim \mathcal{N}(\widetilde{\mu}_v, \widetilde{\Sigma}_v)$

this reason, firms entering the industry before 1973 are excluded from the analysis of firm heterogeneity. Of *all* plants operating in 1996, 75-85 percent were established after 1972, and thus are included in the analysis of firm heterogeneity. These firms account for a similar share of total sales in 1996.

Some "cleaning" of the data was performed. A firm was excluded from the sample if: (i) the value of an endogenous variable is missing for two or more subsequent years; (ii) the firm disappears from the raw data file and then reappears; or (iii) the firm is observed in a single year only. These trimming procedures reduced the data set by 15-20 percent. In addition we removed firms with extreme variations in the endogenous variables, which eliminated an additional 4-8 percent of the observations[11]. Some summary statistics are presented in Table 1.

6 Econometric issues

Our econometric model, presented in Section 3, raises a set of econometric issues that we address in this section. These include: (i) estimation of the structural parameters of the model, (ii) consistency of the parameter estimates in the presence of self-selection, and (iii) calculation of the conditional variance ratio CV for the latent variables. Parts of the discussion are quite technical and some readers may initially wish to proceed to the next section presenting the empirical results.

6.1 Estimation

The main challenge in estimating our econometric model (11) is to obtain a computationally convenient representation of the log-likelihood function and its derivatives. Having achieved that, an efficient quasi-Newton algorithm can be applied to maximize the likelihood function with respect to the unknown parameters: $\boldsymbol{\beta} = (\boldsymbol{\Sigma}_\eta, \boldsymbol{\Sigma}_v, \boldsymbol{\Sigma}_e, \boldsymbol{\theta}_k, \mathbf{d})$, where $\mathbf{d}$ denotes the matrix of time-dummies. A state space representation of the model, combined with a decomposition of the log-likelihood function well known from the EM (Expectation Maximization) algorithm, provides an efficient solution to our estimation problem.

[11] Extreme variation means that the *differenced* variables (on log-scale) have a maximum absolute value that is more than four standard deviations away from the (sector specific) mean maximum absolute values.

The state space representation: In order to obtain a state space representation that is useful for estimation purposes, we start by factorizing the covariance matrices $\mathbf{\Sigma}_\eta$ and $\mathbf{\Sigma}_v$, assuming that these have arbitrary rank r ($r \leq 4$):

$$\mathbf{\Sigma}_\eta = \mathbf{\Gamma}_\eta \mathbf{\Gamma}_\eta' \tag{23}$$

$$\mathbf{\Sigma}_v = \mathbf{\Gamma}_v \mathbf{\Gamma}_v'. \tag{24}$$

Equations (23)-(24) are rank-r decompositions of the two covariance matrices $\mathbf{\Sigma}_\eta$ and $\mathbf{\Sigma}_v$, where $\mathbf{\Gamma}_\eta$ and $\mathbf{\Gamma}_v$ are $4 \times r$ lower triangular matrices (i.e. with zeros above the main diagonal). The matrix factors $\mathbf{\Gamma}_\eta$ and $\mathbf{\Gamma}_v$ are uniquely determined, given positivity of the diagonal elements.

With $\mathbf{\Gamma}_\eta$ and $\mathbf{\Gamma}_v$ defined in (23)-(24), equations (11)-(13) can be restated on the following state space form:

$$\begin{array}{l} \mathbf{y}_{it} = \mathbf{G}\boldsymbol{\alpha}_{it} + \mathbf{d}_t + \boldsymbol{\theta}_k \ln K_{i,t-1} + \mathbf{e}_{it} \\ \boldsymbol{\alpha}_{it} = \mathbf{F}_{it}\, \boldsymbol{\alpha}_{i,t-1} + \boldsymbol{\omega}_{it} \end{array} \qquad t = \tau_i, ..., T_i, \tag{25}$$

where the state vector $\boldsymbol{\alpha}_{it}$ has dimension $2r$, and is determined by the equations:

$$\begin{aligned} \boldsymbol{\alpha}_{i,\tau_i-1} &= \mathbf{0}_{2r} \\ \mathbf{G} &= \left[\begin{array}{cc} \mathbf{\Gamma}_\eta & \mathbf{\Gamma}_v \end{array} \right] \\ \mathbf{F}_{it} &= \begin{cases} \mathbf{0}_{2r\times 2r} & t = \tau_i \\ \mathbf{I}_{2r} & t = \tau_i + 1, ..., T_i \end{cases} \\ \boldsymbol{\omega}_{it} &\sim \begin{cases} \mathcal{IN}\left(\begin{bmatrix} \mathbf{0}_r \\ \mathbf{0}_r \end{bmatrix}, \begin{bmatrix} \mathbf{0}_{r\times r} & \mathbf{0}_{r\times r} \\ \mathbf{0}_{r\times r} & \mathbf{I}_r \end{bmatrix} \right) & t = \tau_i \\ \mathcal{IN}\left(\begin{bmatrix} \mathbf{0}_r \\ \mathbf{0}_r \end{bmatrix}, \begin{bmatrix} \mathbf{I}_r & \mathbf{0}_{r\times r} \\ \mathbf{0}_{r\times r} & \mathbf{0}_{r\times r} \end{bmatrix} \right) & t = \tau_i + 1, ..., T_i. \end{cases} \end{aligned} \tag{26}$$

Notice that $\mathbf{G}\boldsymbol{\alpha}_{it} = \mathbf{a}_{it} + \mathbf{v}_i$, since the first r components of $\boldsymbol{\alpha}_{it}$ are the orthogonal latent factors of $\mathbf{a}_{it}$, normalized to have unit variance, while the last r components of $\boldsymbol{\alpha}_{it}$ are the normalized latent factors of $\mathbf{v}_i$.

The likelihood function and its derivatives: Given the state space representation (25)-(26), it is well known that the log-likelihood function can be evaluated for any given parameter value $\boldsymbol{\beta}$ by using the Kalman filter and smoother (see e.g. Harvey (1989)). Let $\mathbf{y}_{i,\to t} = (\mathbf{y}_{i,\tau_i}, ..., \mathbf{y}_{it})$. Then

$$L(\boldsymbol{\beta}) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{t=\tau_i}^{T_i}\left(\ln\left| \mathbf{G}\mathbf{V}_{it\,|t-1}\mathbf{G}' + \mathbf{\Sigma}_e \right| + \mathbf{R}_{it}{}' \left[\mathbf{G}\mathbf{V}_{it\,|t-1}\mathbf{G}' + \mathbf{\Sigma}_e \right]^{-1} \mathbf{R}_{it} \right)$$

where

$$\begin{aligned}
&\mathbf{V}_{it\,|t-1} = E\{(\boldsymbol{\alpha}_{it} - \mathbf{a}_{it|t-1})(\boldsymbol{\alpha}_{is} - \mathbf{a}_{it-1|T_i-\tau_i+1})' \,|\, \mathbf{y}_{i,\to t-1}\} \\
&\mathbf{a}_{it|t-1} = E\{\boldsymbol{\alpha}_{it} \,|\, \mathbf{y}_{i,\to t-1}\} \\
&\mathbf{R}_{it} = \mathbf{y}_{it} - \mathbf{G}\mathbf{a}_{it|t-1} - \mathbf{d}_t - \boldsymbol{\theta}_k \ln K_{i,t-1}.
\end{aligned} \tag{27}$$

Appendix E explains in detail how the Kalman filter and smoother can be applied to the state space form (25) to evaluate the conditional moments in (27), given $\boldsymbol{\beta}$.

While the evaluation of the likelihood function is straightforward, the main challenge is to obtain analytic expressions for the derivatives of $L(\beta)$. The task of obtaining an analytic form for $\frac{\partial L(\beta)}{\partial \beta}$ may seem prohibitive since $L(\beta)$ indirectly depends on $\boldsymbol{\beta}$ through the Kalman filter recursions[12].

Our solution to the problem is to make a somewhat unusual application of techniques associated with the EM (Expectation Maximization) algorithm – an algorithm originally developed by Dempster, Laird and Rubin (1977), and refined by Meng and Rubin (1993), and others.

Let $f(\mathbf{y}, \boldsymbol{\alpha}; \boldsymbol{\beta})$ be the joint density of the observed variables $\mathbf{y} = \{\mathbf{y}_{it}\}$ and the latent variables $\boldsymbol{\alpha} = \{\boldsymbol{\alpha}_{it}\}$. Furthermore, let $f(\boldsymbol{\alpha}|\, \mathbf{y}; \boldsymbol{\beta})$ be the conditional density of $\boldsymbol{\alpha}$, given $\mathbf{y}$. The maximum likelihood estimator, $\widehat{\boldsymbol{\beta}}$, is the maximum of the log-likelihood $L(\boldsymbol{\beta})$, where

$$L(\boldsymbol{\beta}) = \ln f(\mathbf{y}; \boldsymbol{\beta}). \tag{28}$$

Since

$$f(\mathbf{y}; \boldsymbol{\beta}) = \frac{f(\mathbf{y}, \boldsymbol{\alpha}; \boldsymbol{\beta})}{f(\boldsymbol{\alpha}|\, \mathbf{y}; \boldsymbol{\beta})},$$

(28) can be rewritten as

$$L(\boldsymbol{\beta}) = \ln f(\mathbf{y}, \boldsymbol{\alpha}; \boldsymbol{\beta}) - \ln f(\boldsymbol{\alpha}|\, \mathbf{y}; \boldsymbol{\beta}). \tag{29}$$

Taking the expectation of both sides in (29) with respect to $f(\boldsymbol{\alpha}|\, \mathbf{y}; \boldsymbol{\beta}')$, where $\boldsymbol{\beta}'$ is an arbitrary parameter value, gives:

$$L(\boldsymbol{\beta}) = M(\boldsymbol{\beta}|\boldsymbol{\beta}') - H(\boldsymbol{\beta}|\, \boldsymbol{\beta}'), \tag{30}$$

[12] In principle one could find the derivatives recursively by applying the chain rule to each iterations of the Kalman filter. However, the programming task would be enormous, and even if one were able to obtain the derivatives through a herculean effort, repeated use of the chain rule would magnify round off error due to numerous matrix multiplications and lead to imprecise calculations.

where

$$M(\boldsymbol{\beta}|\boldsymbol{\beta}') = \int \ln f(\mathbf{y}, \boldsymbol{\alpha}; \boldsymbol{\beta})\, f(\boldsymbol{\alpha}|\, \mathbf{y}; \boldsymbol{\beta}')\, d\boldsymbol{\alpha}$$

$$H(\boldsymbol{\beta}|\boldsymbol{\beta}') = \int \ln f(\boldsymbol{\alpha}|\mathbf{y}; \boldsymbol{\beta})\, f(\boldsymbol{\alpha}|\, \mathbf{y}; \boldsymbol{\beta}')\, d\boldsymbol{\alpha}.$$

While the decomposition (30) is not useful in calculating $L(\boldsymbol{\beta})$, it has the following extremely important property:

$$\left.\frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}'} = \left.\frac{\partial M(\boldsymbol{\beta}|\boldsymbol{\beta}')}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}'}, \tag{31}$$

which follows from the fact that $\boldsymbol{\beta}'$ is the maximizer of $H(\boldsymbol{\beta}|\boldsymbol{\beta}')$ (by Kullback's inequality), and hence a stationary point. As shown in Appendix E, the derivatives $\frac{\partial L(\boldsymbol{\beta}')}{\partial \boldsymbol{\beta}}$ can easily be obtained by *analytic* differentiation of $M(\boldsymbol{\beta}|\boldsymbol{\beta}')$. Furthermore, the Hessian of $L(\boldsymbol{\beta})$ at the ML estimate $\widehat{\boldsymbol{\beta}}$ can be obtained by *numerical* differentiation of $\left.\frac{\partial M(\boldsymbol{\beta}|\widehat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}}$, yielding a computationally simple estimator of the covariance matrix of $\widehat{\boldsymbol{\beta}}$.

6.2 Identification, attrition and consistent estimation

Discussing identification of the model (11)-(13) in Section 3.2, we noticed that the question is complicated by entry, and, in particular, sample attrition. We can exploit the results of Cox (1975) and Little and Rubin (1987), which show that a pseudo likelihood function – that is, the likelihood obtained by treating the exit times T_i as if they were fixed indices – yields consistent estimators in the presence of systematic selection, provided the stochastic process, $\mathbf{y}_{it}$, satisfies the so-called missing at random (MAR) condition[13]. The MAR condition needed in our case is (assuming $\tau_i = 1$ for all firms):

$$f(\mathbf{y}_{it}|\chi_{it}, \mathbf{y}_{i1}, .., \mathbf{y}_{i,t-1}; \boldsymbol{\beta}) = f(\mathbf{y}_{it}|\mathbf{y}_{i1}, .., \mathbf{y}_{i,t-1}; \boldsymbol{\beta}),\ t = 1, ..., T \text{ and } i = 1, .., N, \tag{32}$$

where $f(\cdot|\cdot)$ is generic notation for conditional probability density, χ_{it} is the indicator variable, which is 1 if the firm is active in year t, and 0 otherwise, and $\boldsymbol{\beta}$ is the model parameters. As discussed in Raknerud (2001), equation (32) says that information about survival in year t should not help us to predict $\mathbf{y}_{it}$, given the *history* of the observed variables $\mathbf{y}_{i1}, ..., \mathbf{y}_{i,t-1}$.[14] A situation where MAR fails is, say, if the firm knows by the

[13]See Raknerud (2001) for a more in-depth discussion of firm exit and the MAR-condition. Moffitt, Fitzgerald and Gottschalk (1999) refer to the MAR condition as selection on observables.

[14]Notice, however, that the MAR assumption does not exclude firms from having private information that affects their exit decisions, e.g. information about scrap values. See Raknerud (2001).

end of year $t-1$ what its efficiency will be in year t, and chooses to exit if this efficiency is below some threshold. In this case, the value of χ_{it} gives information about $\mathbf{y}_{it}$ not being contained in $\mathbf{y}_{i1}, ..., \mathbf{y}_{i,t-1}$.

Identification of $\boldsymbol{\beta}$ based on the pseudo likelihood function is achieved provided (32) holds and $\boldsymbol{\beta}$ is identified in the model without attrition. This result holds even if exit depends on $\boldsymbol{\beta}$. Thus, we use the term likelihood throughout this paper when, in fact, we consider a pseudo likelihood.

In the presence of self-selection, the MAR assumption is substantially more general than the assumptions required for consistency of widely-used panel data estimators based on the (generalized) method of moments[15].

6.3 Calculation of the conditional variance ratio

The conditional variance ratio (CV), defined in (22), is the ratio of the variances for the unobservables, i.e.

$$CV = \frac{Var\left\{v_i | i \in M_T\right\}}{Var\left\{a_{iT} | i \in M_T\right\}} = \frac{\text{tr } Var\left(\mathbf{v}_i | i \in M_T\right)}{\text{tr } Var\left(\mathbf{a}_{iT} | i \in M_T\right)},$$

where the last equality holds if the structural model is valid. This section explains how $\text{Var}\{\mathbf{v}_i | i \in M_T\}$ and $\text{Var}\{\mathbf{a}_{iT} | i \in M_T\}$ can be estimated.

First note that from (25), $\mathbf{a}_{iT} = \mathbf{G}\mathbf{E}_1\boldsymbol{\alpha}_{iT}$ and $\mathbf{v}_i = \mathbf{G}\mathbf{E}_2\boldsymbol{\alpha}_{iT}$, for selection matrices

$$\mathbf{E}_j = \begin{bmatrix} \delta_{j1}\mathbf{I}_r & \mathbf{0}_{r\times r} \\ \mathbf{0}_{r\times r} & \delta_{j2}\mathbf{I}_r \end{bmatrix}, \quad j = 1, 2,$$

where δ_{jk} is the Kroencker delta function (which is one if $j = k$ and zero otherwise). Hence

$$CV = \frac{\text{tr } Var\left(\boldsymbol{\alpha}_{iT} | i \in M_T\right) \mathbf{E}_2{}'\mathbf{G}'\mathbf{G}\mathbf{E}_2}{\text{tr } Var\left(\boldsymbol{\alpha}_{iT} | i \in M_T\right) \mathbf{E}_1{}'\mathbf{G}'\mathbf{G}\mathbf{E}_1}.$$

From (27) and the rule of iterated expectation:

$$\begin{aligned} &Var\{\boldsymbol{\alpha}_{iT} | i \in M_T\} \\ &= E\{Var\left(\boldsymbol{\alpha}_{iT} | i \in M_T, \mathbf{y}_{i,\to T}\right) | i \in M_T\} + Var\{E\left(\boldsymbol{\alpha}_{iT} | i \in M_T, \mathbf{y}_{i,\to T}\right) | i \in M_T\} \\ &= E\{\mathbf{V}_{iT|T} \; | i \in M_T\} + Var\{\mathbf{a}_{iT|T} \; | i \in M_T\}, \end{aligned}$$

[15] The covariance structure (14) cannot be estimated from sample analogues: If exit is endogenous, $\text{Cov}(\mathbf{y}_{it}, \mathbf{y}_{is} | \max(s,t) \leq T_i)$ will not in general be given by (14) even if MAR holds. Hence the sample covariance matrix ceases to provide consistent estimators for the model parameters. See, however, Abowd, Crepon and Kramarz (2001) who propose a weighted moment estimator that is consistent under the MAR assumption, provided exit probabilities are known or can be estimated.

where the last equality follows from the MAR assumption:

$$f(\boldsymbol{\alpha}_{iT}|i \in M_T, \mathbf{y}_{i,\to T}) = f(\boldsymbol{\alpha}_{iT}|\mathbf{y}_{i,\to T}). \tag{33}$$

Both $E\{\mathbf{V}_{iT|T}|i \in M_T\}$ and $Var\{\mathbf{a}_{iT|T}|i \in M_T\}$ can be estimated from the cross section of firms operating in year T, by the empirical mean and variance of $\mathbf{V}_{iT|T}$ and $\mathbf{a}_{iT|T}$, respectively.

7 Empirical results

This section, which presents our empirical results, is divided into two parts. First, we argue that our structural model presented in Section 3 accounts well for the empirical patterns in most of the industries we consider. On the basis of the structural model, we can construct an estimate of each firm's efficiency every year. The second part of our results explores these estimates. We show that permanent differences dominate differences generated by cumulated, firm-specific innovations in explaining observed firm heterogeneity in all the industries we consider. Finally, we examine the performance of young firms and how selection systematically eliminates firms with low efficiency.

7.1 The validity of our structural model

The results in Tables 2 and 3 largely support our structural model presented in Section 3. Table 2 presents the estimated eigenvalues from the factor decompositions described in Section 3.2. The second column presents the four estimated eigenvalues, $\widehat{\sigma}^2_{\eta,(j)}$, of the covariance matrix for the idiosyncratic innovations, $\boldsymbol{\Sigma}_\eta$. In all the industries, the largest eigenvalue is at least an order of magnitude larger than the second eigenvalue. The same pattern is present in the third column, presenting the four estimated eigenvalues $\widehat{\sigma}^2_{v,(j)}$ of the covariance matrix of the permanent differences, $\boldsymbol{\Sigma}_v$. The largest eigenvalue is also an order of magnitude larger than the second largest eigenvalue in all industries for $\boldsymbol{\Sigma}_v$.

These patterns of eigenvalues show that the persistent differences in performance can largely be summarized by the first latent factors $a_{it,(1)}$ and $v_{i,(1)}$, as they account for at least 90 percent of the variation in $\mathbf{a}_{it}$ and $\mathbf{v}_i$, respectively. This conclusion is confirmed by the last columns in Tables 2 and 3, which present a (pseudo-) R^2-measure varying between .97 and .98 in the four-factor model (Table 2), and between .93 and .96 in the one factor model (Table 3)[16]. Thus, there is only a marginal increase in R^2 when going from the rank-one to the rank-four model. The excellent fit of the model with only one latent factor supports our conclusion that a single permanent component and a single random walk component are largely adequate as a summary of firm performance[17].

As pointed out in Section 3.2, our structural model does not only impose a rank condition on $\mathbf{\Sigma}_\eta$ and $\mathbf{\Sigma}_v$. These matrices should also have the structure that follows from $\boldsymbol{\theta}_a$ (see Section 3 and, in particular, (10) and (15)). That is, the structural model in Section 3 requires that the three first components within each eigenvector should be the same. Furthermore, the eigenvectors of $\mathbf{\Sigma}_\eta$ and $\mathbf{\Sigma}_v$ should be identical (see (19)).

The estimates for the eigenvector in the one-factor model are presented in Table 3, with standard deviations in parentheses. A first look at these results indicates that in four of the six sectors (NACE 29-33), the results for the eigenvector estimates are in good agreement with our structural model. In two industries, Plastics and Transport equipment, our estimates show that the labor variable is less responsive to idiosyncratic innovations than sales and materials, contrary to the prediction by the model in Section 4. The deviation in these two industries may be interpreted as evidence for innovations that are labor-saving or that the technology is non-homothetic (with, roughly speaking, some scale economies for labor). Another explanation could be adjustment costs, but recall that the results in Table 3 refer to responses to persistent changes in efficiency[18].

Formal χ^2-tests of the structural restrictions on the eigenvectors $\mathbf{u}_\eta$ and $\mathbf{u}_v$ are presented in Table 4. While all structural restrictions are clearly rejected in the two industries, Plastics and Transport equipment, the structural hypotheses are largely maintained for the other four sectors. However, in Machinery the restrictions on $\mathbf{u}_v$ (and consequently the hypothesis $\mathbf{u}_\eta = \mathbf{u}_v$) are rejected, despite the fact that the estimates and standard deviations in Table 3 appear to be consistent with the null hypothesis. This outcome

[16] Our pseudo R^2-measure is

$$R^2 = 1 - \frac{\text{tr } \widehat{Var}(\widehat{\mathbf{e}}_{it})}{\text{tr } \widehat{Var}(\mathbf{y}_{it} - \widehat{\mathbf{d}}_{it})},$$

where $\widehat{\mathbf{e}}_{it} = \mathbf{y}_{it} - E(\mathbf{v}_i + \mathbf{a}_{it}|\mathbf{y}_{i,\to T_i}) - \widehat{\boldsymbol{\theta}}_k \ln K_{i,t-1} - \widehat{\mathbf{d}}_t$ (the expectation is evaluated at the estimated parameters and $\widehat{Var}(\cdot)$ denote the sample variance).

[17] A single factor model is an essential, maintained assumption in most empirical studies of firm performance, including Marschak and Andrews (1944) and Olley and Pakes (1996).

[18] Griliches and Hausman (1986) report an elasticity of labor to non-transitory changes in output, which is about the same as the elasticity for materials, while Biørn and Klette (1999) report higher elasticities for materials.

should, however, be interpreted in view of the particularly large number of firms in this sector. As is well known, rejection of any null-hypothesis is only a question of having a sufficiently large data set, since the power of our test tends to one for the slightest departure from the null hypothesis[19]. Machinery is clearly the largest sector (see Table 1), and the rejection of the structural model in this case seems to be due to a very large sample size, rather than substantial evidence that the structural model misrepresents our data.

The eigenvector coefficient in the fourth equation, i.e. the capital accumulation equation, in columns 2 and 3 of Table 3, is small and suggests that the link between innovations and investment is, perhaps, surprisingly weak. However, this is consistent with the capital adjustment model considered in Section 3.1, when the coefficient $\kappa_a = \frac{\delta\alpha}{(1-\alpha)(1-\varepsilon)}$ is small (see (8)). Recall that δ is the the depreciation rate of capital, which is typically a small number ($\approx .05$), while $\alpha \in (0,1)$ reflects adjustment costs.

The coefficients of lagged capital, $\ln K_{i,t-1}$, for each of the four equations in our system (9) are presented in the fourth column in Table 3. The coefficient is slightly less than one in the capital accumulation equation, consistent with moderately decreasing returns to scale.

The last column in Table 2 depicts the four eigenvalues from a decomposition of $\mathbf{\Sigma}_e$, the covariance matrix associated with transient shocks. The results show that the transient shocks are not dominated by a single, common latent factor, in contrast to the persistent shocks. That is, transient fluctuations are not common across the four endogenous variables. We notice that the variance generated by the transient variance component is of the same magnitude as the variance of the innovation component, i.e. $tr\left(\mathbf{\Sigma}_e\right) \approx tr\left(\mathbf{\Sigma}_\eta\right)$. The transient fluctuations account for mean reversion in the dynamic process for the observable variables as pointed out in Section 5.2.

Summarizing our results so far, we conclude that our simple, structural model of firm behavior imposes heavy constraints on the data that are largely fulfilled in at least four of the six industries.

[19]See e.g. Leamer (1983) for a discussion of this issue.

7.2 Permanent differences dominate

Using our estimated model, we can now examine the origin and evolution of differences in efficiency across firms. Table 5 presents various measures of the magnitude of permanent efficiency differences and differences generated by cumulated innovations within each of the six industries. Columns 2 and 3 present the variance in permanent differences and the variance in cumulated innovations. The ratio of these variances, presented in column 4, shows how many years innovations must be cumulated in order to account for as much of the heterogeneity as the permanent differences. These ratios are considerably larger than the average age (column 5) among the firms established after 1972, suggesting that the variance of the permanent efficiency differences accounts for the larger fraction of the non-transient firm heterogeneity in all industries.

These results do not, however, provide a fully satisfactory measure of the importance of permanent differences in explaining the observed variation in firm performance, since they neglect the issue of exit and self-selection. We argued in Section 4 that a better measure is provided by the *conditional* variance ratio, which presents the variance ratio among surviving firms. The conditional variance ratios for each industry in 1996 are presented in column 6. The pattern from the previous columns remains, i.e. the variance of the permanent differences is larger than the variance in the cumulated, idiosyncratic innovations in all industries. The conditional variance ratios vary from 1.2 in Electrical instruments (NACE 31) to 2.6 in Medical instruments (NACE 33) and Transport equipment (NACE 35). In all industries, we find that the conditional variance ratio is at least as large as the unconditional variance ratio. We conclude that in all six industries the permanent differences in efficiency across firms dominate the differences in the cumulated innovations.

7.3 Further results

There is considerable selection that systematically eliminates firms with low efficiency. This can be seen from the ratios in the last column of Table 5. These ratios show that the actual variance in efficiency among surviving firms, accounting for selection, is considerably smaller than the predicted variance in the absence of selection[20].

[20]Similar findings have been presented in a number of studies, as surveyed by Foster, Haltiwanger and Krizan (2001). However, our measurement of efficiency differs from the previous studies. The negative

In all industries there is a strong, *negative* correlation between the permanent efficiency levels v_i and the subsequent innovations, a_{iT} (on average -.40). Our interpretation of this negative correlation is that a firm with a low permanent efficiency level must have a high growth in efficiency in its subsequent years in order to survive and *vice versa*. That is to say, selection is based on the firm's overall efficiency, which is the combination of the permanent efficiency levels and the innovations.

Finally, examining *permanent* differences in efficiency, we find *no* systematic trend across cohorts. Our results reveal no vintage-capital effects where more recent cohorts have higher levels of efficiency. However, we do find that younger firms are more innovative than older firms. That is, there is a negative trend in the mean value of the innovations during the first five to six years of a firm's life time. In addition, young firms have more volatile dynamics than older firms. These results on new firms are consistent with the findings in several other studies surveyed in Caves (1998).

8 Conclusion

This paper examines the large differences across firms in terms of supply and demand for labor, materials and capital. With firm level observations from six manufacturing industries covering 24 years, we showed that almost 95 percent of these differences in supply and factor demands can be accounted for by a single, firm-specific, dynamic factor, which we label efficiency in the light of our structural model. Our structural model of firm behavior is based on a simple production function and price taking behavior, and it explicitly accounts for fully optimizing supply and factor demand.

The structural model enables us to investigate the origin and evolution of the differences in efficiency across firms. The empirical results show that *permanent* differences in efficiency dominate among the firms established within the 24-year period we consider, as they exceed differences in cumulated innovations in efficiency by a factor ranging between 1.2 and 2.6 across the six high-tech industries.

The most striking and controversial result from our analysis is its implications for efficiency measurement. We argue that size is a better indicator of efficiency than labor

correlation between the probability of exit and a firm's productivity level has not been striking in previous studies of Norwegian manufacturing firms. See Møen (1998).

productivity, as long as we also account for the fixity of capital. It is well known that differences in firm size should reflect differences in efficiency, while the serious problem we point out with labor productivity as a measure of efficiency differences seems to have been largely neglected in the literature[21].

Our model suggests that differences in labor productivity should be transitory. This is largely true in our data, but not completely. An important research task is to explain why we observe persistent differences across firms in value added per unit of labor input. Our simple framework suggests that differences in efficiency and capital are not sufficient, and a satisfactory explanation must incorporate a more elaborated model of labor demand. Studies of firm level differences in productivity and labor demand deserve an integrated treatment.

[21] See, however, Bernard, Eaton, Jensen and Kortum (2000) and Klette and Kortum (2002).

References

Abowd, J., B. Crepon and F. Kramarz (2001): Moment estimation with attrition. *Journal of the American Statistical Association*, 96, 1223–31.

Anderson, T. (1984): *An Introduction to Multivariate Analysis*. Wiley Publ. Co. (New York).

Bernard, A., J. Eaton, J. Jensen and S. Kortum (2000): Plants and productivity in international trade. NBER Working paper no. 7688.

Biørn, E. and T. Klette (1999): Errors in variables in panel data: The labour demand response to permanent changes in output. *Scandinavian Journal of Economics*, 101, 379–404.

Blundell, R. and S. Bond (1999): GMM estimation with persistent panel data: An application to production functions. IFS Working Paper No. W99/4.

— (2000): GMM estimation with persistent panel data: An application to production functions. *Econometric Reviews*, 19, 321–40.

Brynjolfsson, E. and L. Hitt (2000): Beyond computation: Information technology, organizational transformation and business performance. *Journal of Economic Perspectives*, 14, 23–48.

Carroll, G. and T. Hannan (2000): *The Demography of Corporations and Industries*. New Jersey: Princeton University Press.

Caves, R. (1998): Industrial organization and new findings on the turnover and mobility of firms. *Journal of Economic Literature*, 36, 1947–82.

Cox, D. (1975): Partial likelihood. *Biometrika*, 62, 269–76.

Dempster, A., N. Laird and D. Rubin (1977): Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, B39, 1–38.

Demsetz, H. (1973): Industry structure, market rivalry, and public policy. *Journal of Law and Economics, 16, 1-10..*

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

Fahrmeir, L. and G. Tutz (1994): *Multivariate Statistical Modelling Based on Generalized Linear Models*. New York: Springer.

Førsund, F. and L. Hjalmarsson (1987): *Analysis of Industrial Structure. A Putty-Clay Approach*. Stockholm: The Industrial Institute for Economics and Social Research.

Foster, L., J. Haltiwanger and C. Krizan (2001): Aggregate productivity growth: Lessons from microeconomic evidence. In C. R. Hulten, E. R. Dean and M. J. Harper (eds.), *New Developments in Productivity Analysis*. Chicago: The University of Chicago Press.

Friedman, M. (1993): Do old fallacies ever die? *Journal of Economic Literature*, 31.

Gibbons, R. (2000): Why organizations are such a mess (and what an economist might do about it). Mimeo, 49 pgs., MIT, Sloan School of Management.

Griliches, Z. and J. Hausman (1986): Errors in variables in panel data. *Journal of Econometrics*, 18, 93–118.

Griliches, Z. and J. Mairesse (1998): Production functions: The search for identification. In S. Strøm (ed.), *The Ragnar Frisch Centennial Symposium*. Cambridge (U.K.): Cambridge University Press.

Harvey, A. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge, U.K.: Cambridge University Press.

Heckman, J. (1991): Identifying the hand of the past: Distinguishing state dependence from heterogeneity. *American Economic Review, Papers and Proceedings*, 81, 75–9.

Holbrook, D., W. Cohen, D. Hounshell and S. Klepper (2000): The nature, sources and consequences of firm differences in the early history of the semiconductor industry. *Strategic Management Journal*, 21, 1017–41.

Hopenhayn, H. (1992): Entry, exit, and firm dynamics in long run equilibrium. *Econometrica*, 60, 1127–50.

Ijiri, Y. and H. Simon (1977): *Skew distributions and the sizes of business firms*. Amsterdam: North Holland Publ.Co.

Johansen, L. (1959): Substitution versus fixed production coefficients in the theory of economic growth: Synthesis. *Econometrica*, 27, 157–76.

Jovanovic, B. (1982): Selection and evolution of industries. *Econometrica*, 50, 649–70.

— (2001): Fitness and age: Review of Carroll and Hannan's "Demography of corporations and industries". *Journal of Economic Literature*, 39, 105–9.

Jovanovic, B. and P. Rousseau (2001): Vintage organizational capital. NBER Working Paper no. 8166.

Kitagawa, S. (1996): *Smoothness Priors Analysis of Time Series*. New York: Springer.

Klepper, S. (1996): Entry, exit, growth and innovation over the product life cycle. *American Economic Review*, 86, 562–83.

Klette, T. and Z. Griliches (2000): Empirical patterns of firm growth and R&D-investment: A quality ladder model interpretation. *Economic Journal*, 110, 363–87.

Klette, T. and S. Kortum (2002): Innovating firms and aggregate innovation. CEPR Working Paper 3248.

Lambson, V. (1992): Competitive profits in the long run. *Review of Economic Studies*, 59, 125–42.

Leamer, E. (1983): Model choice and specification analysis. In Z. Griliches and M. Intriligator (eds.), *Handbook of Econometrics, Volume*, 285–325. Amsterdam: North-Holland.

Little, R. and D. Rubin (1987): *Statistical Analysis with Missing Data*. New York: Wiley.

Lucas, R. (1978): On the size-distribution of firms. *Bell Journal of Economics*, 9, 508–23.

Lutkepohl, H. (1996): *Handbook of Matrices*. New York: John Wiley.

Marschak, J. and W. Andrews (1944): Random simultanous equations and the theory of production. *Econometrica*, 12, 143–205.

Meng, X. and D. Rubin (1993): Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, 80, 267–78.

Milgrom, P. and J. Roberts (1990): The economics of modern manufacturing: Products, technology and organization. *American Economic Review*, 80, 511–28.

Møen, J. (1998): *Productivity Dynamics in Norwegian Manufacturing 1980-1990 - An Analysis Based on Microdata*. Statistics Norway, Reports 21/98. (In Norwegian).

Moffitt, R., J. Fitzgerald and P. Gottschalk (1999): Sample attrition in panel data: The role of selection on observables. *Annales d'Economie et de Statistique*, 55-56, 129–52.

Olley, S. and A. Pakes (1996): The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64, 1263–97.

Pakes, A. and R. Ericson (1998): Empirical implications of alternative models of firm dynamics. *Journal of Economic Theory*, 79, 1–45.

Quah, D. (1993): Galton's fallacy and tests of the convergence hypothesis. *Scandinavian Journal of Economics*, 95.

Raknerud, A. (2001): Identification, estimation and testing in panel data models with attrition: The role of the MAR assumption. Paper presented at Econometric Society European Meeting, 2001.

Stokey, N. and R. Lucas (1989): *Recursive Methods in Economic Dynamics*. Cambridge (U.S.): Harvard Univeristy Press.

Sutton, J. (1997): Gibrat's legacy. *Journal of Economic Literature*, 35, 40–59.

— (1998): *Market Structure and Technology*. Cambridge, Mass.: MIT Press.

Appendix A: Some theoretical ideas on firm heterogeneity

We decompose the persistent differences in firm performance into (i) permanent differences that are established already when the firm enters an industry, and (ii) differences that are generated through subsequent, idiosyncratic innovations that accumulate through the firms' life-time[22]. In this appendix, we briefly review the main ideas in the theoretical literature emphasizing efficiency differences permanent to the firms and differences evolving through innovations that are cumulated, respectively.

The importance of permanent differences in efficiency: Which theoretical models can explain large permanent differences across firms that are introduced already when the firms enter the industry? An old idea is the so-called putty-clay model, emphasizing the irreversible nature of a firm's choice of technology. The classical contribution is Johansen (1959)[23]. The putty-clay literature emphasizes that choices of technology are embodied in the capital, which makes adjustment costly as it requires that the existing capital must be replaced.

Recent case studies of the life cycle of firms suggest that *organizational* capital can be as difficult and costly to adjust as physical capital; see e.g. Holbrook, Cohen, Hounshell and Klepper (2000), Carroll and Hannan (2000), Jovanovic (2001) and Jovanovic and Rousseau (2001). For instance, Holbrook et al. document the development of four of the dominating firms in the early history of the semiconductor industry. Their analysis explains how these firms had a hard time adjusting to the new circumstances as the industry evolved, and eventually all the firms failed and were closed down.

Large costs associated with adjustment of the organizational capital has also been a recurrent theme in studies of the productivity effects of new information technology. Milgrom and Roberts (1990) emphasize that implementing new, IT-based just-in-time production requires simultaneous and costly adjustments in a number of distinct and complementary technological and organizational components in order to be productive. Similar findings have emerged in a number of recent firm level studies examining the (often small) productivity gains from IT-investments; see the survey by Brynjolfsson and Hitt (2000).

That re-adjustments of organizational capital are costly and difficult to implement successfully is not surprising in the light of recent advances in the theory of incentives in firms and organizations. This research has revealed how firms are operated through a complicated system of explicit, formal contracts and informal, relational contracts, and why such a system is costly to adjust and renegotiate; see Gibbons (2000).

Finally, we should mention the study by Jovanovic (1982). His study links differences

[22] In his review of models of firm growth and heterogeneity, Sutton (1997) emphasizes essentially the same distinction, i.e. between models where firm heterogeneity is driven either by "intrinsic efficiency differences" or by "random outcomes emanting from R&D programs". The distinction between intrinsic differences and innovations has also been prominent in labor economics, where the two components are referred to as heterogeneity and state dependence, respectively. See e.g. Heckman (1991).

[23] See Førsund and Hjalmarsson (1987), Lambson (1992) and Jovanovic and Rousseau (2001) for further references to subsequent research.

in firm productivity to differences in the skills of the firms' entrepreneur. The simple and basic idea is that more efficient entrepreneurs command larger firms. This model of firm heterogeneity was introduced by Lucas (1978). It was extended by Jovanovic who introduced entrepreneurial uncertainty about their relative efficiency which is gradually resolved as the entrepreneur learns from the performance of his firm. Jovanovic's model has had considerable empirical success, as it provides an explanation for the high degree of turbulence and high exit rate among young firms. The basic idea that efficiency differences are permanent characteristics embedded in the firms as they are established, is in line with the ideas discussed in this section.

The present study does not aim at discriminating among these various theories which all emphasize the important role of permanent efficiency differences across firms. Instead, this brief survey is provided to remind the reader why differences that are introduced when the firms are born may in principle have a considerable influence on subsequent firm performance.

Firm growth through cumulated innovations: Another line of research has focused on differences in firm performance driven by idiosyncratic and cumulated innovations. The basic idea is that firm performance is driven by firm specific learning, R&D, and innovation, involving significant randomness. This line of ideas emphasizes that a firm's relative efficiency and market share slowly, but gradually *changes* over time.

Early research on firm heterogeneity was stimulated by Gibrat's analysis of the skewed size-distribution of firms, and how such skewed size-distributions can be generated from independent firm growth processes. These growth processes are characterized, according to the so-called Gibrat's law, by firm growth rates that are independent of firm size. Simon and his co-authors developed this line of research in the 1960s and 1970s, by exploring firm evolution through formal modelling of the stochastic processes; see Ijiri and Simon (1977). While this early work paid little attention to optimizing behavior and interactions between firms, Hopenhayn (1992) presents a related study of an industry equilibrium generated by interacting and optimizing firms. Firm growth is driven by exogenous stochastic processes, with exit as an endogenous decision[24].

Gibrat's legacy has recently had a revival, not least due to the work by Sutton (1997, 1998). Sutton shows how persistent differences in firm size and a concentrated market structure tend to emerge in models imposing only mild assumptions on the innovation activities in large versus small firms. His work recognizes the essential role of innovation and R&D in explaining large and persistent differences e.g. in firm sizes, but his model deliberately contains little structure, as he searches for robust patterns which are independent of the detailed model structure. A somewhat more structured model of firm growth through learning and innovation is provided by Ericson and Pakes (1995).

Other recent studies of firm growth emphasizing endogenous learning and innovation, have imposed tight structures on their models in terms of the role of R&D and the nature of the innovation process; see Klepper (1996), Klette and Griliches (2000) and Klette and

[24]Hopenhayn's model accounts for differences in initial conditions, as well as idiosyncratic innovations during the firms' life cycles. Our empirical framework is in large parts consistent with his model of firm evolution.

Kortum (2002). These studies confront stylized facts that have emerged from a large number of empirical studies of R&D, innovation and firm growth.

The common theme across all these models is that firm growth can be considered as stochastic processes, with *idiosyncratic innovations*, and a *high degree of persistence*.

In the rest of this study we examine the relative, quantitative importance of permanent differences on the one hand and cumulated innovations on the other, as sources of persistent firm heterogeneity. Clearly, this is only a first step and subsequent research will aim at discriminating among the theories within each of these line of research.

Appendix B: Initial conditions and non-stationary

In our econometric model we have assumed that $\mathbf{a}_{it}$ is a random walk. However, it might be desirable to generalize the dynamics of the latent process. For example:

$$a_{it} = \phi a_{i,t-1} + \eta_{it} \tag{34}$$

would generalize equation (17), where it was assumed that ϕ equals one. Although our assumption greatly simplifies the interpretation and estimation of our model, and is consistent with Gibrat's law (which has received some support in the empirical literature[25]), the cost is that we might unduly restrict the dynamics of the $\mathbf{y}_{it}$-process.

However, our econometric procedure does not critically depend on the exact value of ϕ, and the main results presented in section 7 would not be seriously affected if ϕ is slightly smaller than one (in line with Blundell and Bond (1999) and Blundell and Bond (2000)). The reason for this is that the distributions of main interest in this paper are the conditional distributions of the latent variables given the observed data (see e.g. the construction of the measure CV in section 6.3). In fact these conditional distributions play the same role in our analysis as the posterior distributions in Bayesian statistics, with equation (34) specifying a common "prior" (i.e. unconditional) distribution. Theory and experience from Bayesian statistics show that inferences based on posterior distributions are generally robust with respect to moderate alternations of the prior distribution (see for example Kitagawa, 1996).

Appendix C: Capital accumulation

A linear, non-stochastic case: The firm's capital accumulation solves the functional equation (see Stokey and Lucas (1989), ch. 5.10):

$$V(K_{t-1}) = \max_{K_t} \{F(K_t, K_{t-1}) + \beta V(K_t)\} \tag{35}$$

where $V(K_{t-1})$ is the value function and $\beta = (1+r)^{-1}$ is the discount factor. Assume that $F(K_t, K_{t-1})$ is increasing and strictly concave in K_t, and homogenous of degree one

[25] The empirical literature suggests that Gibrat's law is valid for large and medium sized firms. The validity of Gibrat's law for smaller firms depends on whether the analysis condition on survival. See Sutton (1997) and Caves (1998) for a discussion and further references.

in (K_t, K_{t-1}). Furthermore, consider the special case:

$$F(K_t, K_{t-1}) = \pi K_{t-1} - K_{t-1}\, qc(K_t/K_{t-1}),$$

where $c(K_t/K_{t-1})$ is continuously differentiable, increasing, and strictly convex, and q is the price per unit of capital. Let $c(1) = \delta$, where δ corresponds to the rate of depreciation. The linear homogeneity of $F(K_t, K_{t-1})$ implies that $V(K_{t-1})$ is linear homogenous in K_{t-1} (see Stokey and Lucas (1989), ch. 5.10), i.e.

$$V(K_{t-1}) = vK_{t-1}. \tag{36}$$

Using (36), the first order condition is

$$\begin{aligned} qc'(K_t/K_{t-1}) &= \beta v \\ \Rightarrow K_t &= K_{t-1}\, g(\beta v/q) \end{aligned} \tag{37}$$

or

$$\ln K_t = \ln K_{t-1} + \ln \left[g\left(\frac{\beta v}{q}\right)\right]$$

The functional form (6) yields:

$$c(K_t/K_{t-1}) = \delta \left(1 + \frac{1}{\delta}\left[\frac{K_t}{K_{t-1}} - 1\right]\right)^{1/\alpha}, \quad \alpha \in (0,1)\,. \tag{38}$$

Given (38), it follows from (37) that $g(x) = 1 - \delta + \delta\,(\alpha x)^{\alpha/(1-\alpha)}$ and

$$K_t = K_{t-1}\left\{1 - \delta + \delta\left[\frac{\alpha\beta v}{q}\right]^{\alpha/(1-\alpha)}\right\}. \tag{39}$$

In a stationary state, $K_{t-1} = K_t$. Thus

$$\frac{\alpha\beta v}{q} = 1 \tag{40}$$

and $F(K_t, K_{t-1}) = (\pi - q\delta)K_{t-1}$. From (35) and (36):

$$\begin{aligned} v &= \pi - q\delta + \beta v \\ &\Rightarrow \frac{\pi - q\delta}{1-\beta} = v = \frac{q}{\alpha\beta}, \end{aligned}$$

where the last equality follows from (40). Rearranging terms, $\pi = q\left(\frac{r}{\alpha} + \delta\right)$, which resembles the well-known formula stating that capital's marginal product, π, equals the Jorgensonian user cost of capital, i.e. $q\,(r+\delta)$. With adjustment costs, $\alpha < 1$ and capital's marginal product, π, exceeds this user cost of capital, as expected.

The stochastic case: In the stochastic case, assuming that (A_t, P^t) is Markovian, with $P^t = (P_t, q_t)$, the firm's investment path can be derived from the following Bellman equation:

$$V(A_t, K_{t-1}, P^t) = \max_{K_t} \left\{ \pi(P_t) A_{it}^{1/(1-\varepsilon)} K_{t-1} - q_t K_{t-1} c(K_t/K_{t-1}) + \beta E \left[V(A_{t+1}, K_t, P^{t+1}) | \Omega_t \right] \right\}, \tag{41}$$

where $V(A_t, K_{t-1}, P^t)$ is the value function and $E[\cdot|\Omega_t]$ is the expectation conditional on the information set Ω_t. Assuming the same functional form as above, the main difference from the preceeding case is that:

$$V(A_t, K_{t-1}, P^t) = \nu(A_t, P^t) K_t,$$

while (39) is replaced by

$$K_t = K_{t-1} \left\{ 1 - \delta + \delta \left[\frac{\alpha \beta v(A_t, P^t)}{q} \right]^{\alpha/(1-\alpha)} \right\}, \tag{42}$$

where

$$v(A_t, P^t) = E \left\{ \nu(A_{t+1}, P^{t+1}) | \Omega_t \right\}.$$

After some calculations, we obtain the following functional equation:

$$\nu(A_t, P^t) = \pi(P_t) A_t^{1/(1-\varepsilon)} + \beta(1-\delta) v(A_t, P^t) + \frac{\delta(1-\alpha)}{\alpha} \left(\frac{\alpha \beta v(A_t, P^t)}{q_t} \right)^{\frac{1}{1-\alpha}}. \tag{43}$$

A linearization of $\left(\frac{\alpha \beta v(A_t, P^t)}{q_t} \right)^{\frac{1}{1-\alpha}}$ around $\frac{\alpha \beta v(A_t, P^t)}{q_t} = 1$ (i.e. $K_{t-1} \simeq K_t$), yields

$$\nu(A_t, P^t) \simeq \pi(P_t) A_t^{1/(1-\varepsilon)} - q_t \delta + \beta E \left\{ \nu(A_{t+1}, P^{t+1}) | \Omega_t \right\}. \tag{44}$$

Furthermore, the expression inside the curly brackets in (42) can be approximated as follows:

$$\begin{aligned} \ln \left\{ 1 + \delta \left(e^{\alpha/(1-\alpha) \ln \left(\frac{\alpha \beta v(A_t, P^t)}{q} \right)} - 1 \right) \right\} &\simeq \ln \left\{ 1 + \frac{\delta \alpha}{1-\alpha} \ln \left(\frac{\alpha \beta v(A_t, P^t)}{q} \right) \right\} \\ &\simeq \frac{\delta \alpha}{1-\alpha} \ln \left(\frac{\alpha \beta v(A_t, P^t)}{q} \right) \end{aligned}$$

Let us consider the solution of (44) in the case where A_t is a geometric random walk, independent of P^t. Assume that $(\pi(P_t), q_t)$ is a martingale. Then $E\{\pi(P_{t+1}) A_{t+1}^{1/(1-\varepsilon)} | \Omega_t\} = \lambda \pi(P_t) A_t^{1/(1-\varepsilon)}$ and $E\{q_{t+1} | \Omega_t\} = q_t$. If a solution to (44) exists,

$$\begin{aligned} \nu(A_t, P^t) &= \frac{\pi(P_t) A_t^{1/(1-\varepsilon)}}{1-\lambda\beta} - \frac{\delta q_t}{1-\beta} \\ v(A_t, P^t) &= \frac{\lambda \pi(P_t) A_t^{1/(1-\varepsilon)}}{1-\lambda\beta} - \frac{\delta q_t}{1-\beta}. \end{aligned}$$

Then (42) can be restated as

$$\ln K_t - \ln K_{t-1} = \kappa_t + \frac{\delta\alpha}{(1-\alpha)(1-\varepsilon)} \ln(A_t) + error,$$

where

$$\kappa_t = \frac{\delta\alpha}{1-\alpha} \ln\left(\frac{r\alpha\beta\lambda\pi(P_t)}{q_t\,(r+\alpha\beta\delta)}\right).$$

By a Taylor expansion, the error term can be written:

$$error = \frac{\delta\alpha}{(1-\alpha)} \frac{1}{1+\frac{r}{\alpha\beta\delta}} (x-1) + O\left((x-1)^2\right)$$

where $x \equiv \alpha\beta v(A_t, P^t)/q_t$. The error term is small relative to the leading term when $K_{t-1} \simeq K_t$ (i.e. $x \simeq 1$) *and* $r/(\alpha\beta\delta)$ is large. The capital accumulation equation is then approximately linear in $\ln A_t$ in the neighborhood of "steady state" when adjustment costs are large and depreciation is slow.

Appendix D: NACE sector codes

25 Manufacture of rubber and plastic products

29 Manufacture of machinery and equipment n.e.c.

31 Manufacture of electrical machinery and apparatus n.e.c.

32 Manufacture of radio, television and communication equipment and apparatus

33 Manufacture of medical, precision and optical instruments, watches and clocks

35 Manufacture of other transport equipment

Appendix E: Computational issues

The Kalman filter and -smoother: We shall now use the state space representation

(25)-(26) to derive the conditional moments (27) by means of the Kalman-filter and -smoother. We first define

$$\mathbf{Q}_{it} = Var\{\boldsymbol{\omega}_{it}\}$$

(see (26)). By modifying the exposition in Fahrmeir and Tutz (1994), p. 264, the filtering recursions can be described by the following algorithm:

$$\begin{aligned}
&\text{Kalman filtering:}\\
&\text{For } i = 1, ..., N:\\
\mathbf{a}_{\tau_i-1|\tau_i-1} &= \mathbf{0}_{2r}\\
\mathbf{V}_{\tau_i-1|\tau_i-1} &= \mathbf{0}_{2r\times 2r}\\
&\text{do for } t = \tau_i, ..., T_i:\\
\mathbf{a}_{it|t-1} &= \mathbf{F}_{it}\mathbf{a}_{i,t-1|t-1}\\
\mathbf{V}_{it|t-1} &= \mathbf{F}_{it}\mathbf{V}_{i,t-1|t-1}\mathbf{F}'_{it} + \mathbf{Q}_{it}\\
\mathbf{Z}_{it} &= \mathbf{y}_{it} - \mathbf{d}_t - \boldsymbol{\gamma}_k \ln K_{i,t-1}\\
\mathbf{K}_{it} &= \mathbf{V}_{it|t-1}\mathbf{G}'[\mathbf{G}\mathbf{V}_{it|t-1}\mathbf{G}' + \boldsymbol{\Sigma}_e]^{-1}\\
\mathbf{a}_{it|t} &= \mathbf{a}_{it|t-1} + \mathbf{K}_{it}(\mathbf{Z}_{it} - \mathbf{G}\mathbf{a}_{it|t-1})\\
\mathbf{V}_{it|t} &= \mathbf{V}_{it|t-1} - \mathbf{K}_{it}\mathbf{G}\mathbf{V}_{it|t-1},
\end{aligned} \tag{45}$$

The conditional expectations $\mathbf{a}_{it|T_i}$ and variances $\mathbf{V}_{it\,|T_i}$ are obtained in subsequent backward smoothing recursions (see Fahrmeir and Tutz (1994), p. 265):

$$\begin{aligned}
&\text{Kalman smoothing:}\\
&\text{For } i = 1, ..., N:\\
&\text{do for } t = T_i, ..., \tau_i + 1:\\
\mathbf{a}_{i,t-1|T_i} &= \mathbf{a}_{i,t-1|t-1} + \mathbf{B}_{it}(\mathbf{a}_{it|T_i} - \mathbf{a}_{it|t-1})\\
\mathbf{V}_{i,t-1|T_i} &= \mathbf{V}_{i,t-1|t-1} + \mathbf{B}_{it}(\mathbf{V}_{it|T_i} - \mathbf{V}_{it|t-1})\mathbf{B}'_{it},
\end{aligned} \tag{46}$$

where

$$\mathbf{B}_{it} = \mathbf{V}_{i,t-1|t-1}\mathbf{F}'_{it}\mathbf{V}^{-1}_{it|t-1}.$$

Derivatives of the log-likelihood function: We shall now show how to obtain analytic derivatives of the log-likelihood function using the relation:

$$\left.\frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\beta=\beta'} = \left.\frac{\partial M(\boldsymbol{\beta}|\boldsymbol{\beta}')}{\partial \boldsymbol{\beta}}\right|_{\beta=\beta'}$$

(see (31)).We first need an expression for

$$\begin{aligned}
M(\boldsymbol{\beta}|\,\boldsymbol{\beta}') = -\frac{1}{2}\sum_{i=1}^{N}\sum_{t=\tau_i}^{T_i}\Big(&\ \ln|\boldsymbol{\Sigma}_e| +\\
&E\left\{(\mathbf{y}_{it} - [\boldsymbol{\Gamma}_\eta\ \boldsymbol{\Gamma}_v]\boldsymbol{\alpha}_{it} - \mathbf{d}_{it} - \boldsymbol{\gamma}_k \ln K_{i,t-1})'\ \boldsymbol{\Sigma}_e^{-1}\,(\mathbf{y}_{it} - [\boldsymbol{\Gamma}_\eta\ \boldsymbol{\Gamma}_v]\boldsymbol{\alpha}_{it} - \mathbf{d}_{it} - \boldsymbol{\gamma}_k \ln K_{i,t-1})\ |\mathbf{y}'_{i,\to T_i};\boldsymbol{\beta}'\right\}\Big),
\end{aligned} \tag{47}$$

where the expectation is evaluated at the parameter value $\boldsymbol{\beta}'$. Standard calculations and (27) yield:

$$M(\boldsymbol{\beta}|\,\boldsymbol{\beta}') = -\frac{1}{2}\sum_{i=1}^{N}\sum_{t=\tau_i}^{T_i}\Big(\ \ln|\boldsymbol{\Sigma}_e|$$
$$+\operatorname{tr}\,\boldsymbol{\Sigma}_e^{-1}\left(\mathbf{y}_{it}-[\boldsymbol{\Gamma}_\eta\ \boldsymbol{\Gamma}_v]\mathbf{a}_{it|T_i}-\mathbf{d}_t-\boldsymbol{\gamma}_k\ln K_{i,t-1}\right)\left(\mathbf{y}_{it}-[\boldsymbol{\Gamma}_\eta\ \boldsymbol{\Gamma}_v]\mathbf{a}_{it|T_i}-\mathbf{d}_t-\boldsymbol{\gamma}_k\ln K_{i,t-1}\right)'$$
$$+\operatorname{tr}\,\boldsymbol{\Sigma}_e^{-1}[\boldsymbol{\Gamma}_\eta\ \boldsymbol{\Gamma}_v]\mathbf{V}_{it|T_i}[\boldsymbol{\Gamma}_\eta\ \boldsymbol{\Gamma}_v]'\Big)\,.$$

In practice, the optimization is performed with respect to the Cholseky factors of $\boldsymbol{\Sigma}_e$ to ensure positive definiteness:

$$\boldsymbol{\Sigma}_e = \boldsymbol{\Gamma}_e\boldsymbol{\Gamma}_e',$$

where $\boldsymbol{\Gamma}_e$ is lower triangular. Hence, in the implementation of the optimization algorithm $\boldsymbol{\beta} = (\boldsymbol{\Gamma}_\eta, \boldsymbol{\Gamma}_v, \boldsymbol{\Gamma}_e, \boldsymbol{\gamma}_k, \mathbf{d})$. Analytic expressions for the derivatives of $M(\boldsymbol{\beta}|\,\boldsymbol{\beta}')$ with respect to the components of $\boldsymbol{\beta}$ are easily available (see Lutkepohl (1996)).

Table 1: **Descriptive statistics**

Sector (NACE)	#Firms: Total vs. 1996	Mean output*	Median output	Lab.prod*
Plastics (25)	242/99	1.77 (2.6)	.74	1.39 (.82)
Machinery (29)	1410/514	1.71 (6.3)	.40	1.37 (.92)
Electrical inst. (31)	377/162	3.30 (11.8)	.61	1.18 (.81)
Radio/TV eq (32)	249/86	4.57 (9.9)	.76	1.04 (.64)
Medical inst. (33)	129/73	2.08 (3.9)	.75	1.51 (.81)
Transp. eq. (35)	818/286	7.03 (23.7)	.99	1.30 (.68)

∗ Standard errors in parentheses. All numbers are in logs.

Table 2: **Estimates of eigenvalues and pseudo R^2 in model with four latent factors**

Sector (NACE)	Eigenvalues of Σ_η (Idiosyncratic innov.)	Eigenvalues of Σ_v (Permanent differences)	Eigenvalues of Σ_e (Noise)	Pseudo R^2
Plastics (25)	(.18, .02, .00, .00)	(3.38, .26, .01, .00)	(.19, .08, .04, .02)	0.97
Machinery (29)	(.24, .02, .00, .00)	(2.00, .20, .00, .00)	(.17, .07, .04, .02)	0.98
Electrical inst.(31)	(.24, .01, .00, .00)	(2.17, .23, .01, .00)	(.15, .07, .02, .02)	0.98
Radio/TV eq.(32)	(.35, .03, .00, .00)	(3.27, .22, .00, .00)	(.27, .07, .04, .02)	0.97
Medical inst. (33)	(.28, .02, .00, .00)	(4.07, .15, .01, .00)	(.15, .07, .02, .01)	0.97
Transp. eq. (35)	(.32, .03, .00, .00)	(5.96, .38, .01, .00)	(.20, .10, .04, .03)	0.98

Table 3: **Estimates of eigenvectors and capital coefficients in model with one latent factor.**

Sector (NACE)	Idiosyn. inn.	Intrinsic dif.	Capital coef.	Pseudo R^2
	Estim. (st.dev.)	Estim. (st.dev.)	Estim. (st.dev.)	
Plastics (25)	.62 (.02)	.59 (.03)	.45 (.17)	0.94
	.73 (.04)	.52 (.10)	.56 (.22)	
	.28 (.12)	.60 (.10)	.32 (.14)	
	.01 (.03)	.02 (.02)	.98 (.02)	
Machinery (29)	.57(.01)	.55 (.01)	.58 (.05)	0.93
	.59 (.01)	.56 (.03)	.62 (.06)	
	.56 (.02)	.61 (.04)	.50 (.05)	
	.00 (.01)	.01 (.01)	.99 (.01)	
Electrical Inst. (31)	.58(.03)	.60(.04)	.65 (.07)	0.96
	.60 (.04)	.60 (.05)	.65 (.08)	
	.54 (.09)	.52 (.10)	.64 (.07)	
	.04 (.02)	.01 (.02)	.99 (.01)	
Radio/TV eq.(32)	.58(.01)	.58(.02)	.44(.11)	0.94
	.61 (.03)	.58 (.04)	.46 (.12)	
	.52 (.05)	.56 (.05)	.43 (.09)	
	.00 (.02)	.03 (.03)	.97 (.03)	
Medical Inst. (33)	.58(.03)	.57(.01)	.31(.15)	0.94
	.61 (.06)	.58 (.03)	.35 (.19)	
	.52 (.07)	.56 (.03)	.29 (.12)	
	.03 (.05)	.01 (.01)	.99 (.04)	
Transp. Eq. (35)	.58(.01)	.58(.03)	.44(.05)	0.95
	.76 (.02)	.61 (.06)	.52 (.06)	
	.29 (.05)	.52 (.07)	.38 (.03)	
	.01 (.01)	.03 (.05)	.97 (.01)	

Table 4: **Testing structural restrictions on eigenvectors**

Sector (NACE)	Restrictions u_η			Restrictions u_v			Joint restictions		
	χ^2	d.f.	P-value	χ^2	d.f.	P-value	χ^2	d.f.	P-value
Plastics (25)	23.07	2	.00	5.2	2	.07	34.43	5	.00
Machinery (29)	3.20	2	.20	30.21	2	.00	44.93	5	.00
Electrical Inst. (31)	.69	2	.70	1.78	2	.41	4.15	5	.52
Radio/TV eq.(32)	5.05	2	.08	.23	2	.89	8.93	5	.11
Medical Inst. (33)	1.00	2	.60	.06	2	.96	2.52	5	.77
Transp. Eq. (35)	105.21	2	.00	18.2	2	.00	131.5	5	.00

Table 5: **Measures of the origins of firm heterogeneity.** The variances of cumulative innovations and intrinsic differences, their ratio, average firm age, conditional variance measure (CV), and actual variance versus predicted variance in the absence of selection .

Sector (NACE)	σ_η^2	σ_v^2	$T^{*}=\frac{\sigma_v^2}{\sigma_\eta^2}$	Avg. age	$\frac{tr\ Var(v_i \vert i \in M_T)}{tr\ Var(a_{it} \vert i \in M_T)}$	$\frac{Var(v_i + a_{iT} \vert i \in M_T)}{\sigma_v^2 + \bar{T}\ \sigma_\eta^2}$
Plastics (25)	0.16	2.27	14.2	7.1	2.3	.38
Machinery (29)	0.20	1.66	8.3	6.9	1.7	.46
Electrical inst. (31)	0.20	1.80	9.0	7.2	1.2	.27
Radio/TV eq.(32)	0.32	3.20	10.0	8.5	2.0	.41
Medical inst. (33)	0.23	3.46	15.0	6.7	2.6	.43
Transp. eq. (35)	0.24	4.25	17.7	8.5	2.6	.71

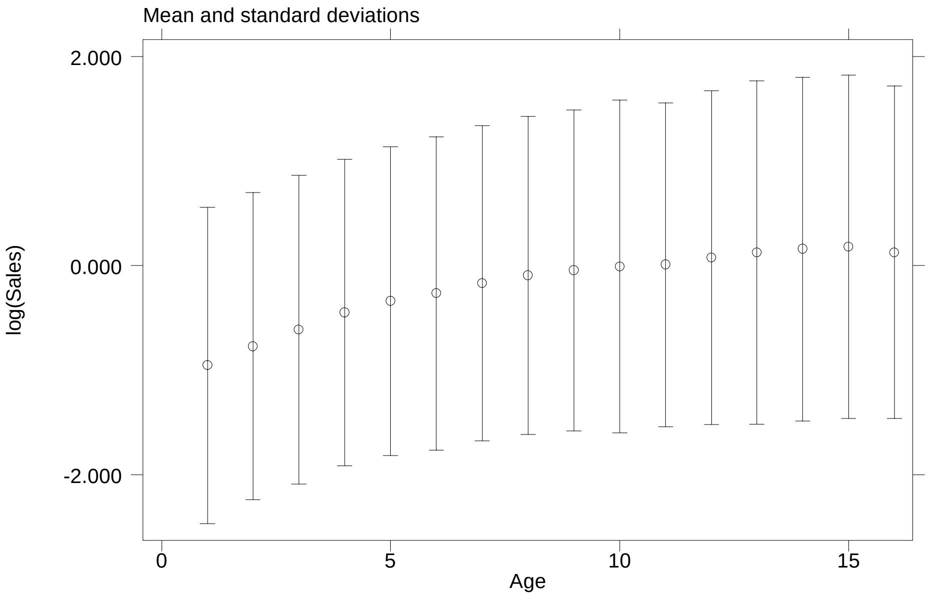

Figure 1: **Differences in log sales as a function of firm age.** Circles indicate the means and whiskers show the standard errors.

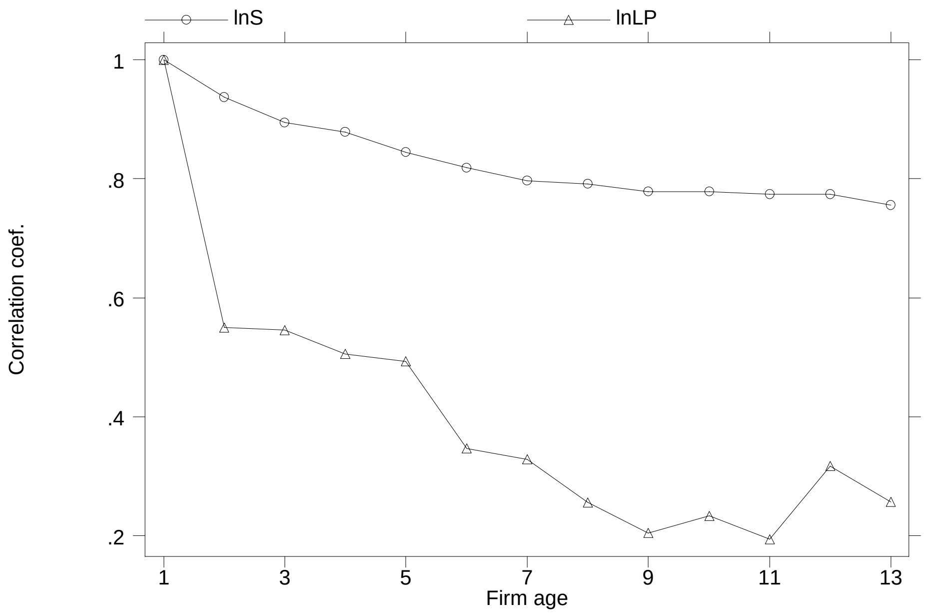

Figure 2: **The correlation between relative performance in a firm's first year and in its subsequent years.** The circles correspond to the correlation coefficents for (log) sales while the triangles refer to (log) labor productivity.

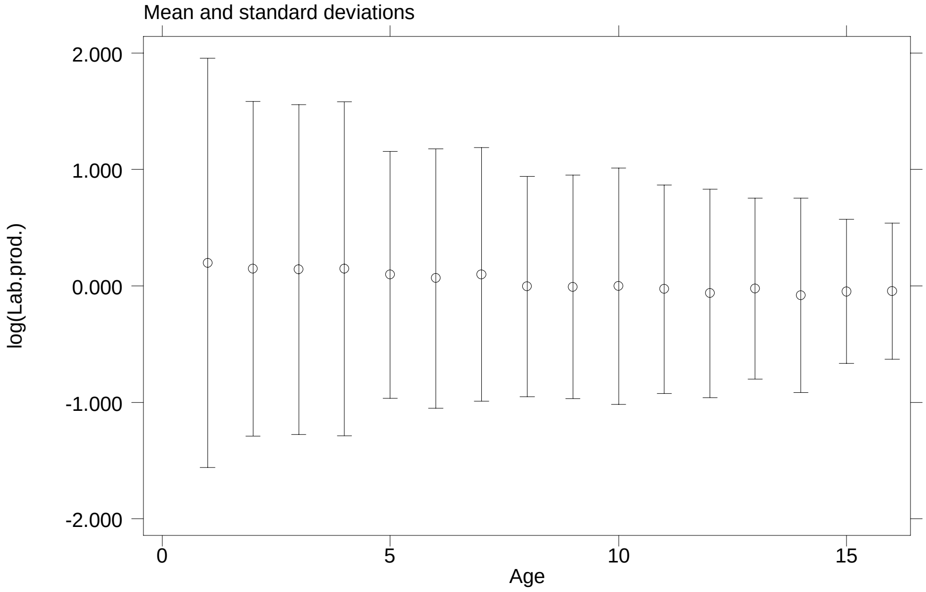

Figure 3: **Differences in log labor productivity as a function of firm age.** Circles indicate the means and whiskers show the standard errors.