How and why do firms differ?*

Tor Jakob Klette† and Arvid Raknerud‡

February 7, 2002

Preliminary draft. Comments are welcome.

ABSTRACT: How do firms differ, and why do they differ even within narrowly defined industries? With evidence from six high-tech, manufacturing industries covering a period of 24 years, 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 light of a structural model. The structural model suggests that this measure is tightly linked to profitability and sales, but unrelated to labor productivity. Our second task is to understand the origin and evolution of the differences in efficiency. Among firms born within the 24 years period we consider, intrinsic (time-invariant) efficiency differences dominate differences generated by firm-specific, cumulated innovations.

JEL classification: xx
Keywords: efficiency, firm heterogeneity, labor productivity, intrinsic differences, firm-specific innovations, state space models,

*This paper has benefitted from comments at seminars at the Institute for Fiscal Studies, the University of Oslo and the University of Helsinki, at the Norwegian School of Economics, the Norwegian School of Management, the Frisch Centre, Statistics Norway, University of Minnesota, and at a NBER Productivity Workshop. 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

There are considerable differences across firms in terms of size, capital intensity, productivity and profitability even within narrowly defined industries, and these differences tend to be highly persistent. With evidence from six manufacturing, industries covering 24 years, we argue that almost 95 percent of the differences in sales, materials, labor, and capital across firms can be accounted for by a single, firm-specific, dynamic factor, which we label efficiency in light of a structural model.[1]

While our model of firm behavior is based on a simple production function and price taking behavior, it explicitly accounts for fully optimizing supply and factor demand. Furthermore, the model accounts for heterogeneity across firms in a flexible way that gives us the opportunity to investigate the origin and evolution of persistent differences in efficiency. In particular, we examine to what extent firms are born with differences in efficiency that are *intrinsic*, i.e. time-invariant, as compared to differences that gradually emerge as the firms evolve. Gradually emerging differences may reflect stochastic, firm-specific (idiosyncratic) *cumulated innovations* as emphasized by Ericson and Pakes (1995), while models emphasizing intrinsic efficiency differences include Jovanovic (1982)[2]. We show that the intrinsic differences in efficiency dominate among the firms born within the 24 years period we consider, as they exceed differences in cumulated innovations by a factor ranging between 1.2 and 2.6 across the six (high-tech) industries.

A large literature on firm heterogeneity has focused on firm performance measured by size (sales or employment), including Pakes and Ericson (1998). Most recent studies of differences in firm performance have, however, focused on differences in efficiency. In competitive environments we expect that 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). We present a structural model highlighting the relationship between size and efficiency. This structural model stresses that differences in firm size are caused by efficiency differences, while emphasizing that the fixity of capital is also essential in explaining differences in firm size.

[1]Pakes (1986) and Lach and Schankerman (1989) have previously exploited dynamic, latent factor models to study differences in firm performance.

[2]Appendix A gives a brief survey of theoretical models emphasizing firm heterogeneity generated by idiosyncratic innovations and intrinsic differences, respectively.

We use the term efficiency rather than productivity, as our structural model suggests that efficiency is unrelated to labor productivity. The argument is simple. Consider firms with different levels of efficiency competing in a frictionless industry[3]. A firm with high efficiency will choose a high level of factor input so that its marginal product is equal to the real wage, which, by assumption, is the same across all firms. With a Cobb-Douglas 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). This argument rises the question of how to make inferences about differences in efficiency from firm level data, which is a central theme of our analysis.

Our econometric framework uses a state space-approach, in combination with the Kalman-filter and -smoother, to decompose the multivariate observations of firm performance in terms of four latent components: (i) firm-specific initial conditions, (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. Using time-series terminology; 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

Are differences in performance across young firms as large as among older firms, or do firms grow more unequal with age? A preliminary answer to this question is suggested in Figures 1-3. Figures 1-3 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. Figure 1 presents the means and standard deviations of log sales as a function of firm age[4]. All observations are measured relative to industry-year means. Not

[3]Assuming diminishing returns for profit-maximization to be well defined.

[4]Graphs for the six individual industries show the same patterns.

surprisingly, the graph shows that young firms are on average substantially smaller than older firms and that firm growth tend to decelerate with age. More interestingly, the graph shows that relative differences in firm size are almost independent of firm age.

Figure 2 shows that the *relative* differences in firm size are highly persistent as the firms get older. That is, the upper graph in Figure 2 displays the correlation coefficient between log sales in the firms' first year and in their subsequent years. The correlation coefficient between log sales in the two first years is 0.94, and it declines steadily to 0.76 when we correlate log sales in the first year and the 12'th year.

These patterns suggest that differences across young firms are as large as among older firms and that the differences are highly persistent, suggesting that firm heterogeneity is generated by intrinsic, i.e. time-invariant, 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 7 years in our sample. Since exiting firms are systematically selected among the least successful firms, we expect the upward trend in Figure 1. Such a systematic selection eliminating the least successful firms should, *cet.par.*, tend to narrow down the differences in firm size, but such narrowing is not visible in Figure 1. At least there must be an offsetting force which tends to make firms grow more unequal with age. Such an offsetting force could be idiosyncratic, cumulated shocks which also explains the declining correlation between a firm's performance in its first year and in its subsequent years in Figure 2.

How should we measure firm performance? Labor productivity is a popular measure. Figure 3 presents means and standard deviations of (log) labor productivity as a function of firm age. We see that the patterns are rather different from that 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 at least equally striking when we turn to Figure 2. The lower graph in Figure 2 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 two first years shows that almost half of the observed variance in labor productivity is temporary fluctuations or noisy data.

[5]Figures 1-3 focus on heterogeneity in new cohorts of firm. In **?**, we show, however, that similar patterns of heterogeneity and autocorrelation are present also 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.

A comparison of the two graphs in Figure 2 rises the question of why differences in size is considerably more persistent than differences in labor productivity. This comparison indicates that labor productivity is a rather noisy measure of efficiency, as we will discuss further below[6].

This 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 intrinsic differences embedded in firms at birth and how the differences evolve over time. It must also account for the considerable noise in the data, self-selection, and yet, it should be flexible enough to let us examine alternative measures of firm performance.

3 A structural model of optimal supply and factor demand

Section 3.1 presents a simple model of optimal supply and factor demand. This model is our basis for making inferences about unobserved differences in efficiency from observations of 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}, A_{it} and $K_{i,t-1}$ denote firm i's output, efficiency and capital in year t, while $F(M_{it}, L_{it})$ is a function aggregating materials and labor inputs. $F(M_{it}, L_{it})$ is homogenous of degree $\varepsilon < 1$.[7] Given prices for output, labor and materials common across firms, $P_t = \left\{p_t, w_t^l, w_t^m\right\}$, and treating $K_{i,t-1}$ as pre-determined, it follows that the short-run cost-function has the following form:

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

[6] See also Bernard, Eaton, Jensen and Kortum (2000) and Klette and Kortum (2001).

[7] Our econometric model is consistent with $F(.)$ varying freely (but preserving the homogeneity property).across years, e.g. allowing for industry-wide, labor augmenting technical change. In order to avoid notational clutter, we have, however, not added a time subscipt to this function and the associated cost function, $C(.)$, presented below.

The set of supply and (short-run) factor demand equations can then be stated

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

According to (3), differences in firm output, material use and labor input, are informative about *unobserved* differences in firm efficiency, conditional on their different capital stocks. The equations in (3) can not, however, be directly exploited to make inferences about the differences in efficiency, as one would expect differences in efficiency to be (positively) correlated with differences in capital. Hence, to obtain an econometric model for inference about differences in efficiency, we must introduce a model of capital accumulation.

The last term, $\mathbf{g}(P_t)$, on the right hand side of (3) is a vector function common across firms which may vary over time as it depends (only) on the common price vector. Its functional form of reflects the properties of the production function (1).

Capital stock dynamics: Consider now the capital stock dynamics derived from each firm's optimal investment behavior. Let $I(K_{it}, K_{i,t-1})$ denote the resources required to change the firm's capital stock from $K_{i,t-1}$ at the end of in period $t-1$ to K_{it} at the end of period t, while p_t^I denotes the price per unit of I. The investment function $I(K_{it}, K_{i,t-1})$ is convex in K_{it} and decreasing in $K_{i,t-1}$. In Appendix C, we examine the firm's investment problem, which can be formulated in terms of dynamic programming

$$\begin{aligned} V(A_{it}, K_{i,t-1}, P^t) \;=\; & \max_{K_{it}} \left\{ \Pi(A_{it}, K_{i,t-1}; P_t) - p_t^I \, I(K_{it}, K_{i,t-1}) \right. \\ & \left. + \, \beta \, E\left[V(A_{i,t+1}, K_{it}, P^{t+1}) | A_{it}, K_{i,t-1} \right] \right\} \end{aligned} \tag{4}$$

where $V(A_{it}, K_{i,t-1}, P^t)$ is the value function, $P^t = \left[P_t, P_{t+1}, p_t^I, p_{t+1}^I\right]$, $E\left[\cdot | A_{it}, K_{i,t-1}\right]$ is the expectation conditional on A_{it} and $K_{i,t-1}$, and β is the discount factor. Appendix C shows that a log-linear approximation to the policy function corresponding to (4), is

$$\ln K_{it} = \kappa_t + \kappa_a \ln A_{it} + \kappa_k \ln K_{i,t-1}. \tag{5}$$

where κ_a is a positive parameter, κ_k is a parameter between zero and one, and κ_t captures changes in the price vector over time. Capital accumulation is, according to (5), driven by cumulated changes in efficiency and changes in input and output prices.

Combining (3) and (5)

$$\mathbf{y}_{it} = \boldsymbol{\theta}_a \ln A_{i1} + \boldsymbol{\theta}_a \ln (A_{it}/A_{i1}) + \boldsymbol{\theta}_k \ln (K_{i,t-1}) + \boldsymbol{\theta}_t. \tag{6}$$

where

$$\begin{aligned} \mathbf{y}_{it} &\equiv \left[\begin{array}{cccc} \ln Q_{it} & \ln M_{it} & \ln L_{it} & \ln K_{it1} \end{array} \right]' \\ \boldsymbol{\theta}_a &= \left[\begin{array}{cccc} \frac{1}{1-\varepsilon}, & \frac{1}{1-\varepsilon}, & \frac{1}{1-\varepsilon}, & \kappa_a \end{array} \right]' \\ \boldsymbol{\theta}_k &= \left[\begin{array}{cccc} \frac{\gamma}{1-\varepsilon}, & \frac{\gamma}{1-\varepsilon}, & \frac{\gamma}{1-\varepsilon}, & \kappa_k \end{array} \right]' \end{aligned} \tag{7}$$

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

The model (6)-(7) suggests that *differences* across firms in the endogenous variables $\mathbf{y}_{it}$ are due to differences in *efficiency* $\boldsymbol{\theta}_a \ln (A_{it})$ and *capital accumulation*, $\boldsymbol{\theta}_k \ln (K_{i,t-1})$. (6) decomposes differences in efficiency into two components: intrinsic differences introduced already when the firms are born, $\boldsymbol{\theta}_a \ln A_{i1}$, and differences in subsequent innovations, i.e. the cumulated changes in efficiency, $\boldsymbol{\theta}_a \ln (A_{it}/A_{i1})$.

Efficiency, profitability and labor productivity: Before we complete the specification of our econometric model by specifying its stochastic properties, we want to address how our model relates differences in efficiency to profitability and labor productivity. According to (3), differences in profitability, Π_{it}, depend on A_{it} and $K_{i,t-1}$:

$$\Pi_{it} = \Pi(A_{it}, K_{i,t-1}; P_t) = \pi(P_t) \left(A_{it} K_{i,t-1}^{\gamma}\right)^{1/(1-\varepsilon)}.$$

where $\pi(P_t)$ is a function only of the input and output prices. On the other hand, (3) suggests 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}. 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, (6)-(7), 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, $\ln(A_{it})$, and, furthermore, that the three first components of the "loading vector" θ_a are equal. Notice, however, that θ_a (and consequently γ) are not identified, because $\ln(A_{it})$ is not observed.

In this section we formulate a more general econometric model which encompasses the structural model. This general econometic model imposes considerably less structure on the data generating process than (6)-(7), and allows us to test the empirical validity of the structural restrictions. Our general model is:

$$\mathbf{y}_{it} = \mathbf{v}_i + \mathbf{a}_{it} + \boldsymbol{\gamma}_k \ln K_{i,t-1} + \mathbf{d}_t + \mathbf{e}_{it}, \quad \tau_i \leq t \leq T, \tag{8}$$

where

$$\mathbf{a}_{it} = \begin{cases} \mathbf{0}_4 & t = \tau_i \\ \mathbf{a}_{i,t-1} + \boldsymbol{\eta}_{it} & t = \tau_i + 1, ..., T \end{cases} \tag{9}$$

($\mathbf{0}_4$ denotes the 4×1 matrix of zeros) and $\mathbf{v}_i, \boldsymbol{\eta}_{it}$ and $\mathbf{e}_{it}$ are 4×1 vectors which have independent, multivariate normal distributions:

$$\mathbf{v}_i \sim \mathcal{IN}(\mathbf{0}_4, \Sigma_v), \quad \boldsymbol{\eta}_{it} \sim \mathcal{IN}(\mathbf{0}_4, \Sigma_\eta), \; \mathbf{e}_{it} \sim \mathcal{IN}(\mathbf{0}_4, \Sigma_e). \tag{10}$$

We have an unbalanced panel data set, where firm i is observed from year $\tau_i \geq 1$ until $T_i \leq T$, where τ_i is the birth date of the firm. The birth dates τ_i have an exogenous distribution, while the exit dates T_i may be endogenous, as we discuss in section 6.2.

When interpreting equation (8) in view of the structural equation (6), 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 measurment errors are captured by $\mathbf{e}_{it}$. It may seem restrictive to assume that $\mathbf{a}_{it}$ is a random walk rather than, say, a mean reverting process. However, as discussed in Appendix B, our econometric procedure does not critically depend on moderate departures from the random walk assumption. In fact, none of our results presented in section 7 would be seriously affected if the $\mathbf{a}_{it}$ process was slightly mean reverting, as suggested by Blundell and Bond (1999, 2000).

The model (8)-(10) encompasses some well-known econometric models of firm heterogeneity as special cases. If $\Sigma_\eta = \mathbf{0}_{4\times4}$, the model corresponds to the fixed effect model widely used to account for firm heterogeneity in the econometric panel data literature. ($\mathbf{0}_{r\times r}$ denotes the $r \times r$ matrix of zeros). When $\Sigma_e = \mathbf{0}_{4\times4}$, 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 Σ_e is a non-zero matrix, the model (8)-(10) implies "mean reversion", in the sense that any component of $\Delta\mathbf{y}_{it}$ will be negatively correlated with the corresponding component of $\mathbf{y}_{it-1}$[8].

Are the parameters of the econometric model (8)-(10) identified? Consider a sample covering two years; $t = 1, 2$. From (8)-(10), ignoring capital for simplicity, we have:

$$\mathrm{Cov}\left(\mathbf{y}_{it}, \mathbf{y}_{is}\right) = \begin{cases} \Sigma_v + \Sigma_\eta \left[\min\left(t, s\right) - 1\right] & t \neq s \\ \Sigma_v + \Sigma_\eta (t-1) + \Sigma_e & t = s. \end{cases} \tag{11}$$

We then obtain: $\mathrm{Cov}(\mathbf{y}_{i2}, \mathbf{y}_{i1}) = \Sigma_v$, $\mathrm{Cov}(\mathbf{y}_{i1}, \mathbf{y}_{i1}) = \Sigma_v + \Sigma_e$, and $\mathrm{Cov}(\mathbf{y}_{i2}, \mathbf{y}_{i2}) = \Sigma_v + \Sigma_\eta + \Sigma_e$. Although identification of the covariance matrices thus appears almost trivial, the situation is complicated by sample attrition, as discussed 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 Σ_v and Σ_η in our general econometric model, while, according to the structural model, these two matrices can be factorized as:

$$\begin{aligned} \Sigma_v &= \boldsymbol{\theta}_a \boldsymbol{\theta}_a' \, Var\left(\ln A_{i1}\right) \\ \Sigma_\eta &= \boldsymbol{\theta}_a \boldsymbol{\theta}_a' \, Var\left[\ln\left(A_{it}/A_{i1}\right)\right]. \end{aligned} \tag{12}$$

If (12) holds, the rank of Σ_η is 1, and all components of $\boldsymbol{\eta}_{it}$ are determined by a single latent factor, say η_{it}:

$$\boldsymbol{\eta}_{it} = \mathbf{u}_\eta \eta_{it},$$

where $\mathbf{u}_\eta$is a 4×1 vector. Hence

$$\mathbf{a}_{it} = \mathbf{u}_\eta a_{it}, \text{ where } a_{it} = \sum_{s \leq t} \eta_{is}. \tag{13}$$

Similarly, we obtain for $\mathbf{v}_i$:

$$\mathbf{v}_i = \mathbf{u}_\eta v_i.$$

[8]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, un-contaminated processes of interest. See also Quah (1993).

As discussed above, the structual parameters $\boldsymbol{\theta}_a$ are not identified. Hence, we normalize $\mathbf{u}_\eta$ such that $\|\mathbf{u}_\eta\| = 1$. Under the structural restrictions, this means that

$$\mathbf{u}_\eta = \frac{\boldsymbol{\theta}_a}{\|\boldsymbol{\theta}_a\|}. \tag{14}$$

It follows that

$$\begin{aligned} v_i &= \|\boldsymbol{\theta}_a\| \ln A_{i1} \\ a_{it} &= \|\boldsymbol{\theta}_a\| \ln \left(A_{it}/A_{i1} \right), \end{aligned} \tag{15}$$

so that $\ln A_{i1}$ and $\ln(A_{it}/A_{i1})$ can be identified up to an (unknown) normalizing constant $\|\boldsymbol{\theta}_a\|$. From the definition of $\boldsymbol{\theta}_a$ in (10), we observe that a testable implication of the structural model is that the three first components of $\mathbf{u}_\eta$will be equal.

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

$$\boldsymbol{\eta}_{it} = \mathbf{u}_{\eta,(1)} \eta_{it,(1)} + ... + \mathbf{u}_{\eta,(r)} \eta_{it,(r)}, \tag{16}$$

where $\mathbf{u}_{\eta,(j)}$ is the eigenvector of Σ_η corresponding to the j'th eigenvalue $\sigma^2_{\eta,(j)}$, with $\|\mathbf{u}_{\eta,(j)}\| = 1$ and $\sigma^2_{\eta,(1)} \geq ... \geq \sigma^2_{\eta,(r)} > 0$. According to our structural model, only the first eigenvalue should be positive while the others should zero, i.e. $\sigma^2_{\eta,(1)} > 0$ while $\sigma^2_{\eta,(j)} = 0$, $j \geq 2$. Hence, if our structural model is correct, the ratio $\sigma^2_{\eta,(1)} / \sum_{j=1}^4 \sigma^2_{\eta,(j)}$ should be very close to one. The denominator in this ratio is equal to $\mathrm{tr}(\Sigma_\eta)$. Hence, this ratio can be interpreted as a pseudo-R^2 measure expressing the fraction of the total variation in the latent factors (in the trace-sense) which are accounted for by the first latent factor. We will also consider a similar decomposition of Σ_v and examine a similar ratio, i.e. $\sigma^2_{v,(1)} / \sum_{j=1}^4 \sigma^2_{v,(j)}$, where $\sigma^2_{v,(j)}$ is the j'th largest eigenvalue.

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

4 Why firms differ in efficiency?

We shall now demostrate how our econometric framework allows us to decompose differences in efficiency and *quantify* the relative importance of intrinsic differences and cumulated innovations, while extracting industry wide changes, as well as transient fluctuations and noise present in the data. Given the validity of our structural model, a natural measure of the importance of intrinsic differences relative to idiosyncratic innovations in a particular year, say T, is

$$V \equiv \frac{Var\{\ln A_{i1}\}}{Var\{\ln(A_{iT}/A_{i1})\}}. \tag{17}$$

Note that V is identified even if $\ln A_{it}$ is not: from (13)-(15) it follows that

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

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.

In our basic econometric model (8), v_i and a_{iT} are uncorrelated, while endogenous exit may cause v_i and a_{iT} to be correlated when we condition on survival. Indeed, our empirical results reveal that v_i and a_{iT} are *negatively* correlated among the surviving firms, as we discuss in section 7.3. These considerations have led us to focus on a modified version of (17): Let M_T be the set of firms that operate in year T. We define the *conditional variance* ratio, CV, as

$$\begin{aligned} CV &\equiv \frac{Var\{\ln A_{i1} | i \in M_T\}}{Var\{\ln(A_{iT}/A_{i1}) \mid i \in M_T\}} \\ &= \frac{Var\{v_i | i \in M_T\}}{Var\{a_{iT} | \ i \in M_T\}}. \end{aligned} \tag{18}$$

In contrast to V, CV takes the endogeneity of firm exit into account.

The measure CV is based on all firms operating in year T. As we shall see in Section 6, it 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 also means that CV is considerably less sensitive to the *a priori* asssumption of an random walk process for a_{it}, as CV is essentially a *semi-parametric* measure. We will return to this issue in Section 6.3, where we also elaborate 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. Separate estimates are presented for 6 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[9]. The unit of observation in our data is an establishment in a given year. For convenience, we have labelled the unit a firm rather than an establishment, which is not misleading in a large majority of cases, since only 10-20 per cent of the establishments belongs to multi-establishment firms in the sectors we consider[10].

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 is 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[11], since v_i is composed of both intrinsic differences and cumulated innovations (up until 1973) and therfore has a different meaning than for firms born after 1972. For this reason, establishments entering the industry before 1973 are excluded from the analysis of firm heterogeneity. Of *all* plants

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

[10] 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 correlation 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.

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

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.

AR: Criterion for entry into Statistics Norway's Annual Manufacturing Census??

Some "cleaning" of the data was performed: A firm was excluded from the sample if either; (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 variation in the endogenous variables, which eliminated an additional 4-8 percent of the observations[12]. 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: (i) Estimation of the structural parameters of the model, (ii) consistency of the parameter estimates in the presence of self-selection when we do not explicitly model the exit-process, and (iii) calculation of the conditional variance ratio CV for the latent variables. Parts of the discussion are quite technical and some readers may, at first, want to proceed to the next section presenting the empirical results.

6.1 Estimation

The main challenge in estimating our econometric model, (8), by maximum likelihood, 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 $\boldsymbol{\beta} = (\Sigma_\eta, \Sigma_v, \Sigma_e, \boldsymbol{\gamma}_k, \mathbf{d})$, where $\mathbf{d}$ denotes the matrix of time dummies. We shall show below that a state space representation of the model, combined with a decomposition of the log-likelihood function well-known from the literature about the EM (Expectation Maximization) algorithm, provides an efficient solution to compute the log-likelihood function and its derivatives.

[12]Extreme variation means that the *differenced* variables (on log-scale) have a maximum absolute value which is more than 4 standard deviations away from the (sector specific) mean maximum absolute values.

The state space representation: In order to obtain a state space representation of the model that is useful for estimation purposes, we factorize the covariance matrices Σ_η and Σ_v, assuming that these have arbitrary rank r ($r \leq 4$):

$$\Sigma_\eta = \Gamma_\eta \Gamma'_\eta, \tag{19}$$

$$\Sigma_v = \Gamma_v \Gamma'_v. \tag{20}$$

(19)-(20) present the rank-r decompositions of the two covariance matrices Σ_η and Σ_v,.where Γ_η and Γ_v are unique $4 \times r$ lower triangular matrices.

With Γ_η and Γ_v defined by (19)-(20), equations (8)-(10) can be restated on the following state space form:

$$\begin{array}{l} \mathbf{y}_{it} = \mathbf{G}\boldsymbol{\alpha}_{it} + \mathbf{d}_t + \boldsymbol{\gamma}_k \ln K_{i,t-1} + \mathbf{e}_{it} \\ \boldsymbol{\alpha}_{it} = \mathbf{F}_{it}\, \boldsymbol{\alpha}_{i,t-1} + \boldsymbol{\omega}_{it} \end{array} \qquad t = \tau_i, ..., T_i \tag{21}$$

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

$$\begin{aligned} \boldsymbol{\alpha}_{i,\tau_i-1} &= \mathbf{0}_{2r} \\ \mathbf{G} &= \left[\begin{array}{cc} \Gamma_\eta & \Gamma_v \end{array}\right] \\ \mathbf{F}_{it} &= \begin{cases} \mathbf{0}_{2r\times 2r} & t = \tau_i \\ \mathbf{I}_{2r} & t = \tau_i + 1, ..., T_i \end{cases} \\ \boldsymbol{\omega}_{it} &\sim \begin{cases} \mathcal{IN}\left(\begin{bmatrix} \mathbf{0}_r \\ \mathbf{0}_r \end{bmatrix}, \begin{bmatrix} \mathbf{0}_{r\times r} & \mathbf{0}_{r\times r} \\ \mathbf{0}_{r\times r} & \mathbf{I}_r \end{bmatrix} \right) & t = \tau_i \\ \mathcal{IN}\left(\begin{bmatrix} \mathbf{0}_r \\ \mathbf{0}_r \end{bmatrix}, \begin{bmatrix} \mathbf{I}_r & \mathbf{0}_{r\times r} \\ \mathbf{0}_{r\times r} & \mathbf{0}_{r\times r} \end{bmatrix} \right) & t = \tau_i + 1, ..., T_i \end{cases} \end{aligned} \tag{22}$$

Notice that $\mathbf{G}\boldsymbol{\alpha}_{it} = \mathbf{a}_{it} + \mathbf{v}_i$, since the first r components of $\boldsymbol{\alpha}_{it}$ are the 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 (21)-(22), the log-likelihood function can readily be evaluated for any given parameter value β. Let $\mathbf{y}_{i,\to t} = (\mathbf{y}_{i,\tau_i}, ..., \mathbf{y}_{it})$, and define conditional means and variances

$$\begin{aligned} \mathbf{a}_{it|t-1} &= E\{\boldsymbol{\alpha}_{it} \,|\, \mathbf{y}_{i,\to t-1}; \boldsymbol{\beta}\} \\ \mathbf{V}_{it\,|t-1} &= E\{(\boldsymbol{\alpha}_{it} - \mathbf{a}_{it|t-1})(\boldsymbol{\alpha}_{is} - \mathbf{a}_{it-1|T_i-\tau_i+1})' \,|\, \mathbf{y}_{i,\to t-1}; \boldsymbol{\beta}\}. \end{aligned} \tag{23}$$

As explained e.g. in Harvey (1989), the Kalman filter and -smoother can be applied to the state space form to evaluate (23). Furthermore, the log-likeliohood function $L(\boldsymbol{\beta})$ can be calculated as follows:

$$L(\boldsymbol{\beta}) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{t=\tau_i}^{T_i}\Big(\ln|\mathbf{V}_{it\,|t-1}| + \mathbf{R}_{it}{}'\mathbf{V}_{it\,|t-1}^{-1}\mathbf{R}_{it}\Big)$$

where

$$\mathbf{R}_{it} = \mathbf{y}_{it} - \mathbf{G}\mathbf{a}_{it|t-1} - \mathbf{d}_t - \boldsymbol{\gamma}_k \ln K_{i,t-1}.$$

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 $\mathbf{V}_{it\,|t-1}$ and $\mathbf{a}_{it|t-1}$, considered as functions of $\boldsymbol{\beta}$, do not have closed form expressions, but are defined through the Kalman-filter recursions (see Appendix E)[13].

Our solution to the problem is to make an unorthodox application of techiques associated with the EM (Expectation Maximization) algorithm – an algorithm originally developed by Dempster, Laird and Rubin (1977), and refined by and others.

Let $f(\mathbf{y}, \boldsymbol{\alpha}; \boldsymbol{\beta})$ be the joint density of the 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

$$L(\boldsymbol{\beta}) = \ln f(\mathbf{y}; \boldsymbol{\beta}). \tag{24}$$

Since

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

(24) can be rewritten as

$$L(\boldsymbol{\beta}) = \ln f(\mathbf{y}, \boldsymbol{\alpha}; \boldsymbol{\beta}) - \ln f(\boldsymbol{\alpha}|\, \mathbf{y}; \boldsymbol{\beta}). \tag{25}$$

Taking the expectation of both sides in (25) with respect to $f(\boldsymbol{\alpha}|\, \mathbf{y}; \boldsymbol{\beta}')$, where $\boldsymbol{\beta}'$ is an arbitrary parameter value, gives:

$$L(\boldsymbol{\beta}) = M(\boldsymbol{\beta}|\boldsymbol{\beta}') - H(\boldsymbol{\beta}|\, \boldsymbol{\beta}'), \tag{26}$$

[13]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 an herculean effort, repeated use of the chain rule would magnify round off error due to numerous matrix multiplications and lead to inprecise calculations.

where

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

The most important property of the equation (26), for our purpose, is the following:

$$\left.\frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\beta=\beta'} = \left.\frac{\partial M(\boldsymbol{\beta}|\boldsymbol{\beta}')}{\partial \boldsymbol{\beta}}\right|_{\beta=\beta'}, \tag{27}$$

which follows from (26) and the fact that $\boldsymbol{\beta}'$ is the maximizer of $H(\boldsymbol{\beta}|\boldsymbol{\beta}')$ (by Kullback's inequality), and hence a stationary point. The derivatives $\frac{\partial L(\beta')}{\partial \beta}$ can therefore be obtained by *analytic* differentiation of $M(\boldsymbol{\beta}|\boldsymbol{\beta}')$ (see Appendix E). Furthermore, the Hessian of $L(\boldsymbol{\beta})$ at the ML estimate $\widehat{\boldsymbol{\beta}}$ can be obtained by *numerical* differentiation of $\left.\frac{\partial M(\beta|\widehat{\beta})}{\partial \beta}\right|_{\beta=\widehat{\beta}}$, 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 (8)-(10) in section 3.2, we noticed that the question is complicated by entry, and, in particular, sample attrition. However, we can exploit the results of *AR: Cox-ref?* **?** and Little and Rubin (1987), showing 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[14]. The MAR condition needed in our case is:

$$f(\mathbf{y}_{it}|\chi_{it}, \mathbf{y}_{i1}, .., \mathbf{y}_{i,t-1}; \boldsymbol{\beta}) = f(\mathbf{y}_{it}|\mathbf{y}_{i1}, .., \mathbf{y}_{i,t-1}; \boldsymbol{\beta}),\ t = 1, ..., T \text{ and } i = 1, .., N, \tag{28}$$

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. Equation (28) says that information about survival in year t should not help us to predict $\mathbf{y}_{it}$, given the *history* of the observed variables $\mathbf{y}_{i1}, ..., \mathbf{y}_{i,t-1}$.[15] A situation where MAR fails is, say, if the firm anticipated by the end of year $t-1$ what the firm's

[14]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.

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

efficiency will be in year t, and chooses to exit if this anticipated 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 are achieved provided (28) holds and $\boldsymbol{\beta}$ is identified in the model without attrition. This result holds even if exit depends on $\boldsymbol{\beta}$, as discussed in Raknerud (2001). We use the term likelihood throughout this paper when, in fact, we consider a pseudo likelihood.

Notice that, 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[16].

6.3 Calculation of the conditional variance ratio

The conditional variance ratio (CV), defined in (18), is the ratio of the variances for the unobservables, i.e.

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

This section explains how $\mathrm{Var}\{\mathbf{v}_i | i \in M_T\}$ and $\mathrm{Var}\{\mathbf{a}_{iT} | i \in M_T\}$ can be estimated.

First note that from (21), $\mathbf{a}_{iT} = \mathbf{G}\mathbf{E}_1 \boldsymbol{\alpha}_{iT}$ and $\mathbf{v}_i = \mathbf{G}\mathbf{E}_2 \boldsymbol{\alpha}_{iT}$, for selection matrices

$$\mathbf{E}_j = \left[\begin{array}{cc} \delta_{j,1} \mathbf{I}_r & \mathbf{0}_{r \times r} \\ \mathbf{0}_{r \times r} & \delta_{j,2} \mathbf{I}_r \end{array} \right], \quad j = 1, 2,$$

where $\delta_{j,k}$ is the indicator function which is one if $j = k$ and zero otherwise. Hence

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

From (23) and the rule of iterated expectation:

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

[16]Note that the covariance structure (11) cannot be estimated from sample analogues: If exit is endogenous, $\mathrm{Cov}(s_{it}, s_{is} | \max(s,t) \leq T_i)$ will not in general be given by (11) (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 etimator which is consistent under the MAR assumption, provided exit probabilities are known or can be estimated.

where the last equality follows from the MAR assumption:

$$f(\boldsymbol{\alpha}_{iT}|i \in M_T, \mathbf{y}_{i,\to T}) = f(\boldsymbol{\alpha}_{iT}|\mathbf{y}_{i,\to T}). \tag{29}$$

Both $E\{\mathbf{V}_{iT|T}|i \in M_T\}$ and $Var\{\boldsymbol{\alpha}_{iT|T}|i \in M_T\}$ 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 $\boldsymbol{\alpha}_{iT|T}$, respectively:

$$E\{\mathbf{V}_{iT|T}|i \in M_T\} \approx \frac{1}{N_T}\sum_{i\in M_T}\mathbf{V}_{iT|T}$$

$$Var\{\mathbf{a}_{iT|T}\ |i \in M_T\} \approx \left(\frac{1}{N_T}\sum_{i\in M_T}\mathbf{a}_{iT|T}\mathbf{a}_{iT|T}{}'\right) - \left(\frac{1}{N_T}\sum_{i\in M_T}\mathbf{a}_{iT|T}\right)\left(\frac{1}{N_T}\sum_{i\in M_T}\mathbf{a}_{iT|T}\right)'.$$

where N_T is the number of firms in the set M_T.

7 Empirical results

This section presents our empirical results, which can be divided into two parts. First, we argue that our structural model presented in section 3 accounts for the empirical patterns in most of the industries we consider. With reference to the structural model, we can construct an estimate of each firm's efficiency every year. The second part of our results shows that intrinsic 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 Table 2 and 3 support our structural model presented in section 3, in most of the industries. Table 2 presents the estimated eigenvalues from the factor decompositions described in section 3.2. The second column presents the four eigenvalues, $\sigma^2_{\eta,(j)}$, of the covariance matrix for the idiosyncratic innovations, Σ_η. In all the industries, the largest eigenvalue is at least an order of magnitude larger than the second. The same pattern is present in the third column, presenting the four eigenvalues $\sigma^2_{v,(j)}$ of the covariance matrix of the intrinsic differences, Σ_v. The largest eigenvalue is an order of magnitude larger than the second largest eigenvalue in all industries also for Σ_v.

These patterns of eigenvalues show that the persistent differences in performance can largely be summarized by the first latent factors $v_{i,(1)}$ and $a_{it,(1)}$, as they account for at least 90 percent of the variation in $\mathbf{v}_i$ and $\mathbf{a}_{it}$, respectively. This conclusion is confirmed by the last column in Tables 2 and 3, which present a (pseudo-) R^2 varying between $.97 - .98$ in the four-factor model (Table 2), and between $.93 - .96$ in the one factor model (Table 3)[17]. The excellent fit of the model with one latent factor supports our conclusion that a single, latent, time-invariant component and a single, latent, random walk component, is largely adequate as a summary of firm performance[18].

As pointed out in section 3.2, our structural model does not only impose a rank condition on Σ_η and Σ_v. These matrices should in addition have the structure that follows from $\boldsymbol{\theta}_a$ (see section 3.2 and, in particular, equation 14). That is, the structural model in section 4 suggests that the three first components of the loading vector should be the same; both for the idiosyncratic innovations and for the intrinsic differences. The validity of the restrictions on the eigenvectors can be considered from Table 3, presenting estimates of the factor loadings in a one factor model. In all but two industries, the parameter estimates are consistent with the restrictions on the factor loadings imposed by our structural model.

In two industries, Plastics and Transport equipments, 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 persistent changes in efficiency[19].

[17] 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 $\widehat{\mathbf{e}}_{it} = \mathbf{y}_{it} - E(\mathbf{v}_i + \mathbf{a}_{it}|$ all the data on firm $i) - \widehat{\boldsymbol{\gamma}}_k k_{i,t-1} - \widehat{\mathbf{d}}_t$ (the expectation is evaluated at the estimated parameters and $\widehat{Var}(\cdot)$ denote the sample variance).

[18] 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).

[19] 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.

The fourth factor loading in column 2 and 3, i.e. corresponding to the capital variable equation, is small and suggests that the link between innovations and investment is, perhaps surprisingly, weak. Such a weak link may reflect a more complicated capital adjustment pattern than considered in section 3.1, due to e.g. non-convex adjustment costs. The capital coefficient for each of the four equations in our system, (6), are presented in the fourth column in Table 3. We notice that in two industries discussed above, Plastics and Transport equipments, the capital coefficient in the labor equation is smaller than the capital coefficients in the equations for sales and materials. The fourth coefficient is close to one, suggesting that the capital process is almost an independent random walk.

The last column in Table 2 depicts the four eigenvalues from a decomposition of Σ_e, the covariance matrix associated with transient shocks. The results show that the transient shocks are not dominated by a single, common latent factor, i.e. 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. $tr\left(\Sigma_e\right) \approx tr\left(\Sigma_\eta\right)$. The transient fluctuations account for mean reversion in the dynamic process for the observable variables (see footnote 5).

7.2 Intrinsic differences dominate

Table 4 presents various measures of the magnitude of intrinsic differences and differences generated by cumulated innovations within each of the six industries. Column 2 and 3 present the variance in intrinsic differences and the variance in cumulated innovations. The ratio of these variances, presented in column 4, shows how many years innovations must be cumulated to account for as much of firm heterogeneity as the intrinsic differences. These ratios are considerably larger than the average age among the firms established after 1972, suggesting that the variance of the intrinsic efficiency differences accounts for the larger fraction of the non-transient firm heterogeneity in all industries.

These results do not, however, provide a satisfactory measure of the importance of intrinsic 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 for the 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. that the variance of the intrinsic 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 the intrinsic differences in efficiency dominate the differences in the cumulated innovations in all six industries.

7.3 Younger firms are more innovative

Several studies have suggested that younger firms are more innovative than older firms, even when controlling for selection-bias; see Caves (1998). Our results is consistent with this evidence, as seen in the upper chart in Figure 4 presenting the mean value of the innovations as a function of firm age together with 95 percent point-wise confidence intervals[20]. Both graphs in Figure 4 presents the average pattern across industries[21]. From the upper chart in Figure 4, we see that for one year old firms the innovation mean is about 0.3, gradually decreasing to 0 after four to five years, and then stabilizing around -0.1 after about eight years. The negative trend in the mean value of the innovations is clearly significant during at least the first five years of a firm's life time.

The lower chart in Figure 4 shows that the variance of the innovations also declines with age. The innovation variance is 1.5 for the youngest firms and then steadily decreases, eventually stabilizing around 1 after ten years.

We get a clear impression from Figure 4 that new firms have more volatile and turbulent dynamics than older firms. On average, younger firms have better innovations than older firms, but younger firms are also more likely to experience less favorable innovations (large negative η_{it}-values).

[20]Recall that, given the MAR-condition, our estimation procedure is not biased by self-selection.

[21]The innovations from different industries have been rescaled with each industry's variance of the innovations, to highlight the (common) shape of the age profile. The details of the estimation procedure is presented in Appendix E.

7.4 Systematic selection

Figure 5 presents the estimated cohort-specific means of the intrinsic differences for all sectors combined, together with confidence intervals[22]. The graph shows *no* systematic trend where more recent cohorts have higher intrinsic efficiency levels, as suggested by vintage-capital models[23].

The last column in Table 4 shows that among the surviving firms, there is a strong, *negative* correlation between the intrinsic efficiency levels v_i and the subsequent innovations, a_{iT}. Recall that, according to our model, the intrinsic differences and the cumulated innovations are uncorrelated in the population, i.e. in the absence of sample selection. Our interpretation of this negative correlation is that a firm with a low intrinsic efficiency level must have a high growth in efficiency its subsequent years in order to survive and *vice versa*. That is, selection is based on the firm's overall efficiency which is the combination of the intrinsic efficiency levels and the innovations.

Figure 6 compares the actual variance, accounting for selection, with the predicted variance in the absence of selection across all the industries. The actual variance is considerably smaller than the predicted variance. This shows that selection reduces differences in efficiency by systematically eliminating firms with low efficiency. Similar findings have been presented in a number of studies, as surveyed by Foster, Haltiwanger and Krizan (2001)[24]. However, our measurement of efficiency differs from the previous studies.

8 Conclusions

(To be rewritten:)

How do firms differ, and why do they differ even within narrowly defined industries? We show that the non-transient 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 light of a structural model. The structural model suggests that this

[22]As in Figure 2, we have rescaled the initial conditions from different industries by each industry's variance for the initial conditions to highlight the (common) shape of the age profile. The details of the estimation procedure is presented in Appendix E.

[23]The classical contribution is Johansen (1959). See also Førsund and Hjalmarsson (1987), Lambson (1992) and Jovanovic and Rousseau (2001).

[24]The negative correlation between the probability of exit and a firm's efficiency level has not been striking in previous study of Norwegian manufacturing firms. See Møen (1998).

measure is tightly linked to profitability and sales, but unrelated to labor productivity. Our second task has been to explain the origin and evolution of the persistent differences in efficiency. We find that among firms born within a period of 24 years, intrinsic (time-invariant) efficiency differences dominate differences generated by firm-specific, cumulated innovations. Our results also confirm previous findings suggesting that young firms are more innovative and have more volatile and turbulent dynamics than older firms. Finally, we show that selection systematically eliminates low-performing firms.

To be explored:

- Labor productivity, second factor, $\theta^l < \theta^i, i = q, m$
- Innovation survey, η_{it} and ΔLP

Our results highlight the rigidity of organizations, and suggests that competition does not eliminate inefficiencies within firms, but rather that competition promote efficiency by eliminating inefficiencient firms. Similar observations have recently been made by Geroski (2000): "[T]he rise and fall of organizations is likely to be driven by selection pressures rather than by adaption...[O]rganizations are rather rigid and do not change easily to market forces." This is not a new perspective[25], but it is steadily reinforced as new evidence based on newly available firm level data is accumulated. However, our paper has emphasized that much of this research should pay further attention to the problem of measuring efficiency and performance at the micro level. We have presented a somewhat new approach, and we will elaborate the relationship between our performance measure, traditional measures of productivity and other, outside evidence of innovation and efficiency in future research.

[25] It has for a long time been advocated by evolutionary economists, see Winter (1987).

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.

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.

Geroski, P. (2000): Exploring the niche overlaps between organizational ecology and industrial economics. CEPR Working Paper no. 2649.

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.

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.

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 (2001): Innovating firms and aggregate innovation. Mimeo. University of Oslo and Boston University.

Lach, S. and M. Schankerman (1989): Dynamics of R&D and investment in the scientific sector. *Journal of Political Economy*, 97, 880–904.

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

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. (1986): On patents, R&D, and the stock market rate of return. *Journal of Political Economy*, 96, 390–409.

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.

Winter, S. (1987): Competition and selection. In J. Eatwell, M. Milgate and P. Newman (eds.), *The New Palgrave: A Dictionary of Economics*. New York: MacMillan Press.

Appendix A: Some theoretical ideas on firm heterogeneity

We decompose the persistent differences in firm performance into (i) intrinsic 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[26]. The time-invariant part is referred to as the intrinsic differences, while the cumulated part is labelled idiosyncratic, cumulated innovations (or just cumulated innovations). In this appendix, we briefly review the main ideas in the theoretical literature emphasizing efficiency differences intrinsic to the firms and differences evolving through innovations that are cumulated, respectively.

The importance of intrinsic differences in efficiency: Which theoretical models can explain large intrinsic 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)[27]. 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

[26] 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).

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

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 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 intrinsic 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[28].

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

[28]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.

of the innovation process; see Klepper (1996), Klette and Griliches (2000) and ?. 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 intrinsic 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

Tba.

Appendix C: Capital accumulation

As stated above, the firm's investment problem can be formulated in terms of dynamic programming

$$V(A_t, K_{t-1}, P^t) = \max_{K_t} \left\{ \Pi(A_t, K_{t-1}, P_t) - p_t^I I(K_t, K_{t-1}) + \beta E\left[V(A_{t+1}, K_t, P^{t+1}) | A_t \right] \right\}$$

where $V(A_t, K_{t-1}, P^t)$ is the value function, $P^t = \left\{ \left(P_t, p_t^I\right), \left(P_{t+1}, p_{t+1}^I\right) .... \right\}$ and $E\left[\cdot|A_t\right]$ is the expectation conditional on A_t. A_t follows an independent Markov process. Following Stokey and Lucas (1989), ch. 10.4, assume that the investment costs are such that

$$I(K_{t,}K_{t-1}) = K_{t-1}\ c(K_t/K_{t-1})$$

where the function $c(K_t/K_{t-1})$ is zero when its argument is $1-\delta$ or smaller, and continuously differentiable, increasing and strictly convex when its argument (strictly) exceeds $1-\delta$. δ corresponds to the rate of depreciation, which is less than one. If K_{t-1} is sufficiently large, the optimal level of investment is zero, and $K_t = (1-\delta)\, K_{t-1}$. The threshold level for K_{t-1} for which this occurs, $\bar{K}$, is an increasing function of the state variable A_t. On the other hand, if $K_{t-1} < \bar{K}\,(A_t)$, the optimal level of capital accumulation is determined from the first order condition

$$c'(K_t/K_{t-1}) = \frac{\partial}{\partial K_t} \beta E\left[V(A_{t+1}, K_t, P^{t+1}) | A_t \right]. \tag{30}$$

This equation gives a relationship between the optimal level of K_t conditional on K_{t-1} and A_t, i.e. the policy function $K_t = g(K_{t-1}, A_t)$. Lemma 9.5 in Stokey and Lucas (1989) states that when $\Pi(A_t^*, K_{t-1}) - I(K_{t,}K_{t-1})$ is increasing in A_t and K_{t-1}, $E\left[V(A_{t+1}, K_t)|A_t\right]$ is also increasing in A_t and K_{t-1}. It follows that the policy function is increasing in both arguments. (*Restate??*)

Hence, we have that

$$K_t = \begin{cases} (1-\delta)\, K_{t-1} & \text{if } K_{t-1} \geq \bar{K}(A_t) \\ g(K_{t-1}, A_t) & \text{otherwise} \end{cases}$$

where $g(K_{t-1}, A_t)$ is increasing in both arguments and concave in its first argument. A log-linear approximation to this policy function is

$$\ln K_{it} = \kappa_t + \kappa_a \ln A_{it} + \kappa_k \ln K_{i,t-1}. \tag{31}$$

where κ_a and κ_k are both positive. We have added a firm-subscript, i. The constant term has a time-subscript to capture that the capital accumulation will be affected by prices which can vary over time.

We can charactirze κ_k somewhat further by Lemma 9.5 in Stokey and Lucas (1989). This lemma states that when $\Pi(A_t^*, K_{t-1}) - I(K_{t,}K_{t-1})$ is strictly concave in K_{t-1}, $E\left[V(A_{t+1}, K_t)|A_t\right]$ is strictly concave in K_t. Furthermore, the policy function is concave in its first argument, which can be verified as follows. Define $\hat{k}_t = dK_t/K_t$, it follows from (30) that

$$\frac{\hat{k}_t}{\hat{k}_{t-1}} = \frac{c''}{c'' - m}$$

which is positive and below one since $c'' > 0$ and m is negative:

$$m \equiv K_t \frac{\partial^2}{\partial K_t^2} \beta E\left[V(A_{t+1}, K_t, P^{t+1})|A_t\right] < 0.$$

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

(21)-(22) to derive the conditional moments (23) by means of the Kalman-filter and -smoother. We first define

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

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

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

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):

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

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 derivatives of the log-likelihood function using the relation:

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

We first need an expression for

$$
\begin{aligned}
M(\boldsymbol{\beta}|\,\boldsymbol{\beta}') = -\frac{1}{2}\sum_{i=1}^{N}\sum_{t=\tau_i}^{T_i}\Big(&\ \ln|\Sigma_e| + \\
&E\left\{(\mathbf{y}_{it} - [\Gamma_\eta\ \Gamma_v]\boldsymbol{\alpha}_{it} - \mathbf{d}_{it} - \boldsymbol{\gamma}_k \ln K_{i,t-1})'\ \Sigma_e^{-1}\left(\mathbf{y}_{it} - [\Gamma_\eta\ \Gamma_v]\boldsymbol{\alpha}_{it} - \mathbf{d}_{it} - \boldsymbol{\gamma}_k \ln K_{i,t-1}\right)\,|\mathbf{y}'_{i,\to T_i};\boldsymbol{\beta}'\right\}\Big),
\end{aligned}
\tag{32}
$$

where the expectation is evaluated at the parameter value $\boldsymbol{\beta}'$.

Because (32) is the expectation of a function which is quadratic in $(\boldsymbol{\alpha}_{i,\tau_i}, ..., \boldsymbol{\alpha}_{i,T_i})$, to evaluate $M(\boldsymbol{\beta}|\,\boldsymbol{\beta}')$ we only need to calculate the conditional moments:

$$\begin{aligned} \mathbf{a}_{it|T_i} &= E\{\boldsymbol{\alpha}_{it}\,|\,\mathbf{y}_{i,\rightarrow T_i};\boldsymbol{\beta}^{')}\} \\ \mathbf{V}_{it\,|T_i} &= E\{(\boldsymbol{\alpha}_{it}-\mathbf{a}_{it|T_i})(\boldsymbol{\alpha}_{it}-\mathbf{a}_{it|T_i})'\,|\,\mathbf{y}_{i,\rightarrow T_i};\boldsymbol{\beta}'\}. \end{aligned} \tag{33}$$

which are available from the Kalman filter and -smoother. Standard calculations yield:

$$\begin{aligned} M(\boldsymbol{\beta}|\,\boldsymbol{\beta}') &= -\frac{1}{2}\sum_{i=1}^{N}\sum_{t=\tau_i}^{T_i}\Big(\ \ln|\Sigma_e| \\ &+ \mathrm{tr}\ \Sigma_e^{-1}\left(\mathbf{y}_{it}-[\Gamma_\eta\ \Gamma_v]\mathbf{a}_{it|T_i}-\mathbf{d}_t-\boldsymbol{\gamma}_k\ln K_{i,t-1}\right)\left(\mathbf{y}_{it}-[\Gamma_\eta\ \Gamma_v]\mathbf{a}_{it|T_i}-\mathbf{d}_t-\boldsymbol{\gamma}_k\ln K_{i,t-1}\right)' \\ &+ \mathrm{tr}\ \Sigma_e^{-1}[\Gamma_\eta\ \Gamma_v]\mathbf{V}_{it|T_i}[\Gamma_\eta\ \Gamma_v]'\Big)\,. \end{aligned}$$

In practice, the optimization is performed with respect to the Cholseky factors of Σ_e to ensure positive definiteness:

$$\Sigma_e = \Gamma_e\Gamma_e',$$

where Γ_e is lower triangular. Hence, in the implementation of the optimization algorithm $\boldsymbol{\beta} = (\Gamma_\eta, \Gamma_v, \Gamma_e, \boldsymbol{\gamma}_k, \mathbf{d})$. 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)).

Appendix F: Estimating the distribution of age and cohort profiles

This appendix presents some details concerning the estimation of the graphs in Figures 2 and 3. Starting with Figure 2, let $\eta_{it} = \eta_{i,(1)}/\sigma_{\eta,(1)}$ be the *standardized* innovations in the one-factor model, where $\eta_{i,(1)}$ and $\sigma_{\eta,(1)}$ are defined in (16)-(**??**), and define $\overline{\eta}_s \equiv E\left\{\eta_{i,\tau_i+s}\right\}$ – the expected innovation of a firm of age s. Then, by the rule of iterated expectation, we can estimate $\overline{\eta}_s$ using the sample analogy method:

$$\begin{aligned} \overline{\eta}_s &= E\left\{E\left\{\eta_{it}|y_{i,\rightarrow T_i}\right\}\right\} \\ &\approx \frac{1}{n_s}\sum_{i\in N_s}E\left\{\eta_{it}|y_{i,\rightarrow T_i}\right\}, \end{aligned}$$

where $N_s = \{i : \tau_i + s \leq T_i\}$ and n_s are the number of firms in this set.

For fixed s, $Var\left\{\eta_{i,\tau_i+s}\right\} = Var\left\{E\left\{\eta_{i,\tau_i+s}|y_{i,\rightarrow T_i}\right\}\right\} + E\left\{Var\left\{\eta_{i,\tau_i+s}|y_{i,\rightarrow T_i}\right\}\right\}$. Hence, the variance of η_{i,τ_i+s} as a function of the age s, can be estimated from the cross section of the $E\left\{\eta_{i,\tau_i+s}|y_{i,\rightarrow T_i}\right\}$ and the $Var\left\{\eta_{i,\tau_i+s}|y_{i,\rightarrow T_i}\right\}$, which are outputs from the Kalman-smoother (cf. section 6.1 and appendix E).

The relation between intrinsic differences and selection in section 7.3 is studied from the smoothed innovations $\widehat{\mathbf{v}}_i = E\left\{\mathbf{v}_i|y_{i,\rightarrow T_i}\right\}$. Since Σ_v can be well approximated by a rank one matrix, it is enough to study the standardized univariate innovation $v_i \equiv v_{i,(1)}/\sigma_{i,(1)}$, where $v_{i,(1)}$ and $\sigma_{i,(1)}$ are defined in the factor decomposition (**??**)-(**??**).

Table 1: **Descriptive statistics**

Sector (NACE)	**#Firms**	**# Firms in 95**	**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 in model with four latent factors**

Sector (NACE)	Eigenvalues of Σ_η (Idiosyncratic innov.)	Eigenvalues of Σ_v (Intrinsic differences)	Eigenvalues of Σ_e (Noise)	Pseudo
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 factor loadings in model with one latent factor.** St.dev.s are approximately .03, .05, and .06 for the three first components in column 2, 3, and 4, respectively.

Sector (NACE)	Idiosyn. innov.	Intrinsic differences	Capital coef.	Pseudo R^2
Plastics (25)	(.25, .30, .11, .01)	(.90, .78, .92, .03)	(.45, .56, .32, .98)	0.94
Machinery (29)	(.26, .26, .25, .00)	(.71, .73, .79, .02)	(.58, .62, .50, .99)	0.93
Electrical Inst. (31)	(.27, .27, .25, .00)	(.81, .81, .70, .01)	(.66, .66, .64, .99)	0.96
Radio/TV eq.(32)	(.33, .35, .30, .02)	(1.04, 1.05, 1.01, .06)	(.22, .25, .19, .93)	0.94
Medical Inst. (33)	(.28, .30, .25, .02)	(1.08, 1.08, 1.06, .04)	(.31, .36, .26, .99)	0.94
Transp. Eq. (35)	(.28, .37, .14, .01)	(1.20, 1.00, 1.35, .07)	(.45, .53, .38, .98)	0.95

Table 4: **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 correlations between a_{it} and v_i.

Sector (NACE)	σ_η^2	σ_v^2	$T^* = \frac{\sigma_v^2}{\sigma_\eta^2}$	**Avg. age**	$\frac{tr\ Var(v_i \mid i \in M_T)}{tr\ Var(a_{it} \mid i \in M_T)}$	$\rho(a_{it}, v_i)$
Plastics (25)	0.16	2.27	14.2	7.1	2.3	−.27
Machinery (29)	0.20	1.66	8.3	6.9	1.7	−.39
Electrical inst. (31)	0.20	1.80	9.0	7.2	1.2	−.55
Radio/TV eq.(32)	0.32	3.20	10.0	8.5	2.0	−.44
Medical inst. (33)	0.23	3.46	15.0	6.7	2.6	−.50
Transp. eq. (35)	0.24	4.25	17.7	8.5	2.6	−.34

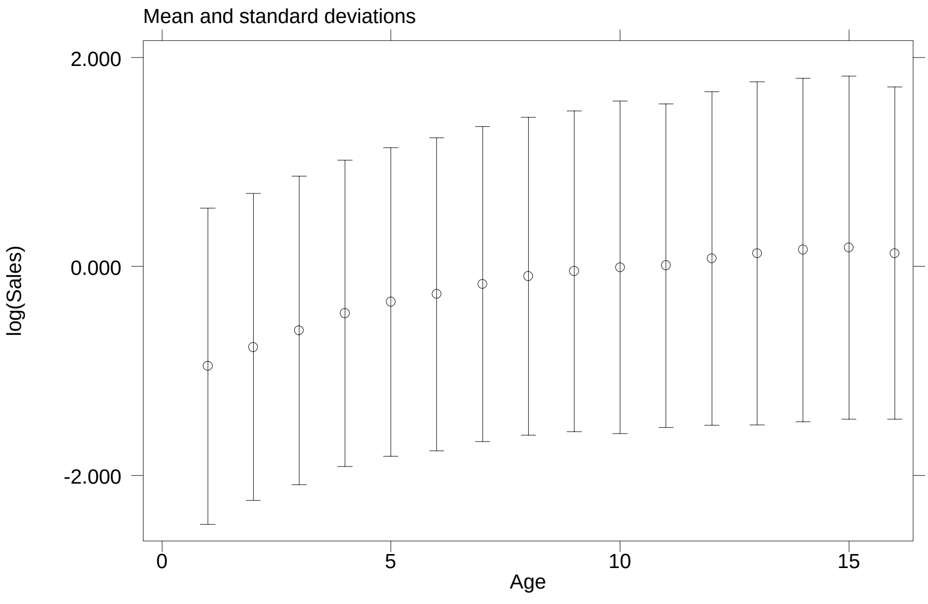


Figure 1: **Differences in log sales as a function of firm age.** Circles indicate the means and whiskers show the standard errors.

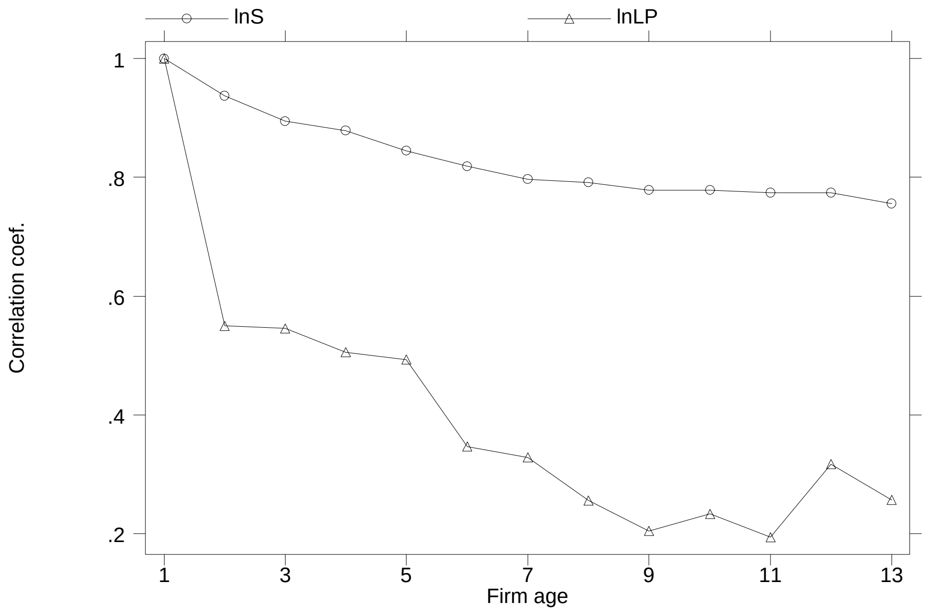


Figure 2: **The correlation between relative performance in a firm's first year and in its subsequent years.** The circles correspond to the correlation coefficents for (log) sales while the triangles refer to (log) labor productivity.

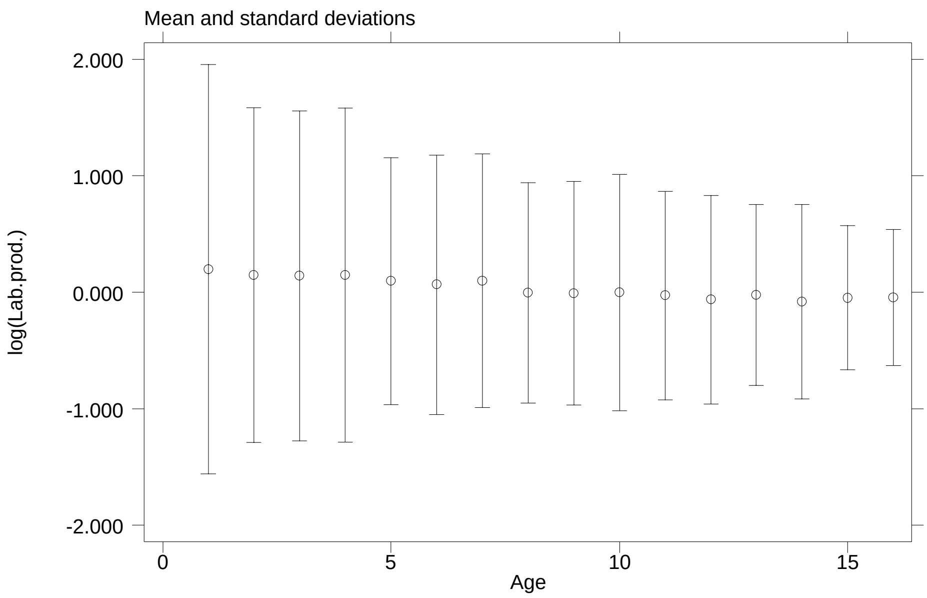
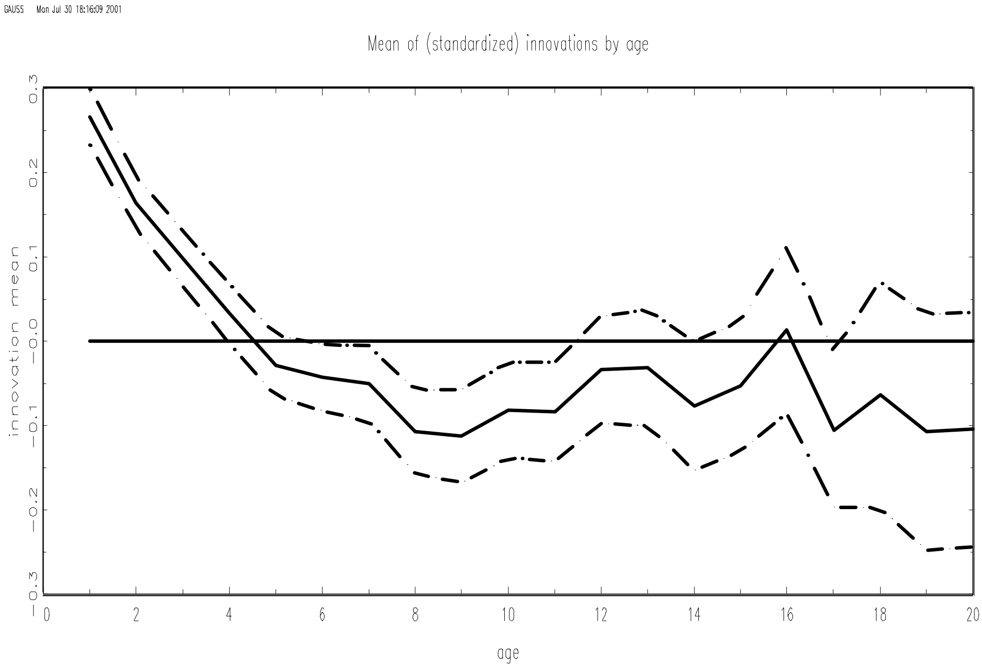


Figure 3: **Differences in log labor productivity as a function of firm age.** Circles indicate the means and whiskers show the standard errors.



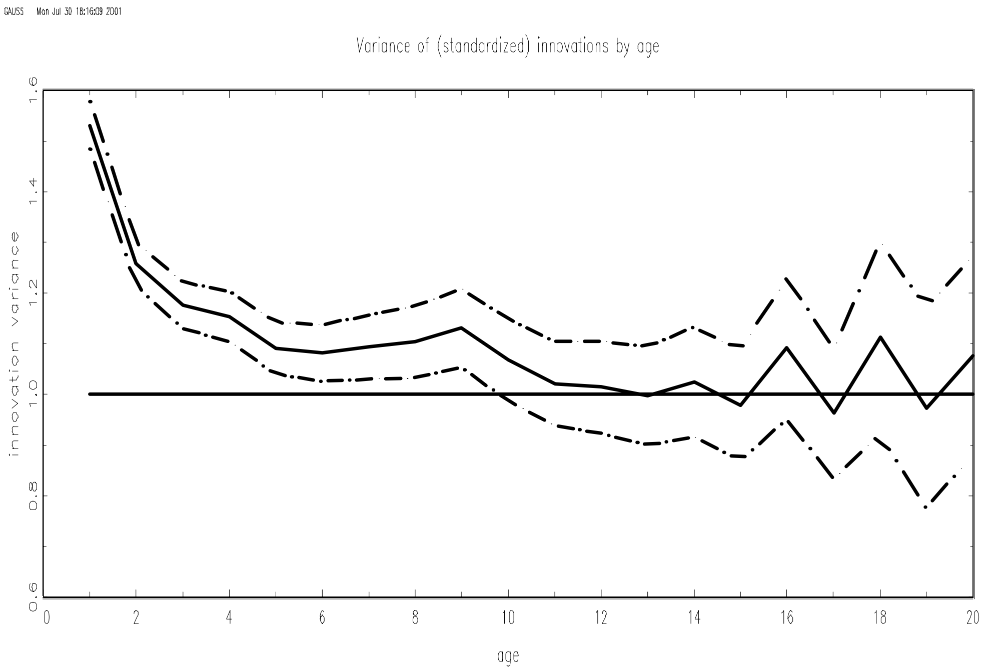
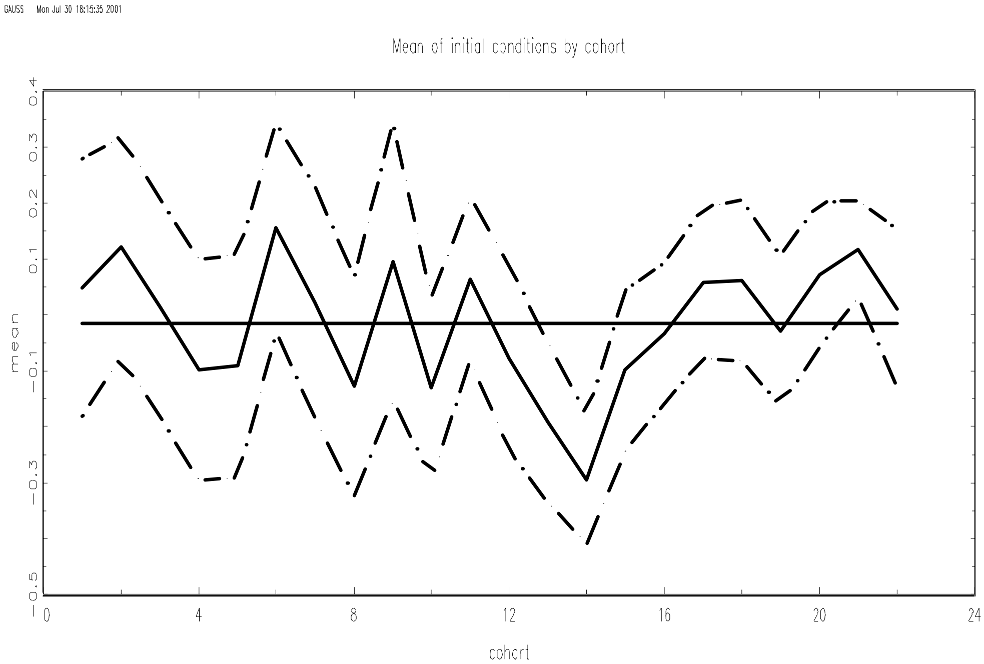


Figure 4: **The mean (upper chart) and variance (lower chart) of the innovations decrease with the age of the firms.** Standard deviations indicated by dotted lines.



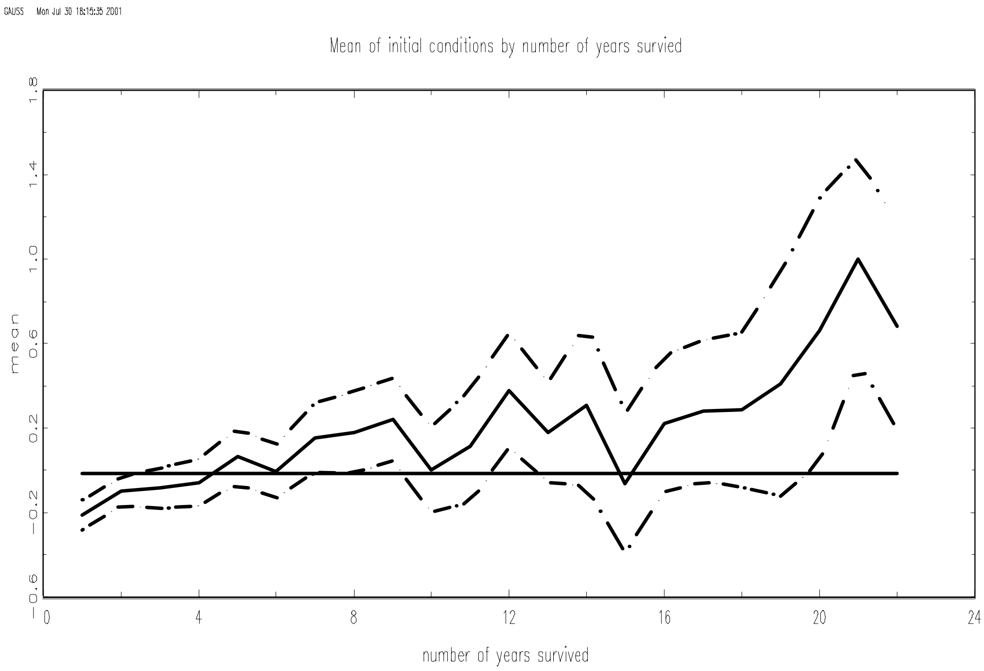


Figure 5: **No systematic differences across cohorts in initial conditions (upper chart). Firms with higher initial productivity live longer (lower chart).** Standard deviations indicated by dotted lines.

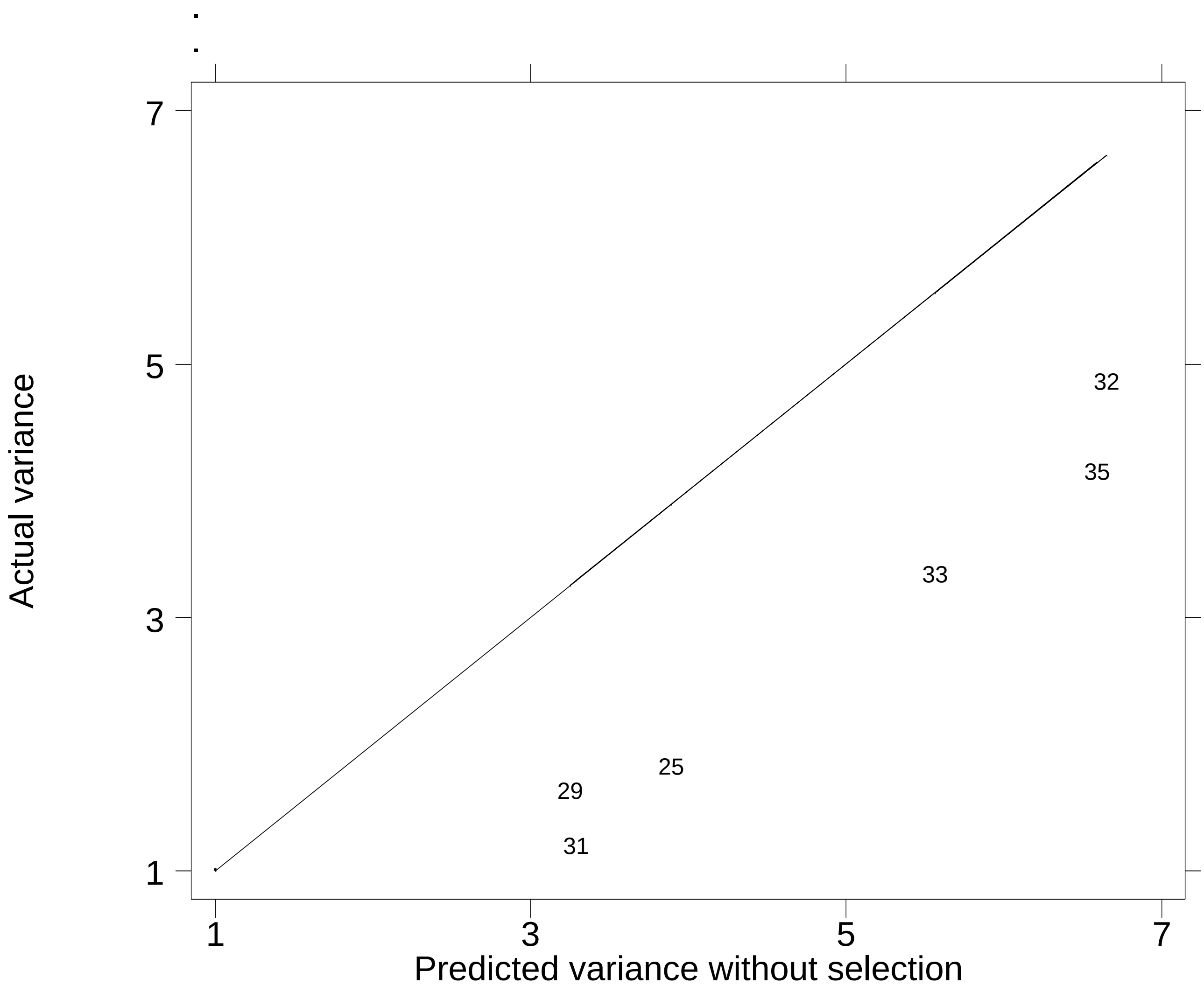


Figure 6: **Observed efficiency differences much smaller than predicted efficiency differences in the absence of selection.** Variances of the observed efficiency differences on the vertical axis and predicted efficiency differences in the absence of selection on the horizontal axis. 45-degree line also presented. Numbers refer to NACE codes for the individual industries (see Appendix C).