How and why do firms differ?

Tor Jakob Klette*
(Department of Economics, University of Oslo and CEPR)

Arvid Raknerud
(Research Department, Statistics Norway)

December 5, 2001

ABSTRACT: How do firms differ, and why do they differ even within narrowly defined industries? We show that 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 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 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 conclusions are based on evidence from six high-tech, manufacturing industries.

Keywords: efficiency, labor productivity, firm heterogeneity, intrinsic differences, innovations, self-selection, unbalanced panel data, state space model
JEL: xx

*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 and at Statistics Norway. 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").

1 Introduction

A number of researchers have reported considerable differences across firms in terms of size, capital intensity, productivity and profitability even within narrowly defined industries, and that these differences are highly persistent over many years. In this study we examine the nature and evolution of these *persistent* differences in performance. First, we examine to what extent various measures of performance can be summarized and related through a model of firm behavior. We show that the non-transient differences in sales, materials, labor, 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. Our conclusion is based on evidence from six manufacturing, high-tech industries over a period of 24 years.

Our second task is to understand the origin and evolution of persistent differences in efficiency in light of two classes of models that allow for heterogeneity among firms. That is, we examine to what extent firms are born with differences in efficiency that are intrinsic and time-invariant, as compared to differences which slowly, but gradually, emerges as firms evolve, e.g. through stochastic, idiosyncratic (or firm-specific) innovations. Models emphasizing intrinsic efficiency differences include Jovanovic (1982), while Ericson and Pakes (1995) present a model of firm heterogeneity driven by stochastic, idiosyncratic innovations emerging from the firms' R&D-activities. We show that the intrinsic and time-invariant differences in efficiency dominate among the firms born within the 24 years period we consider, as they exceeds differences in cumulated innovations by a factor ranging between 1.2 and 2.6 across the six industries.

A large literature on firm growth measure performance by firm size (sales or employment), e.g. 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 dominate or at least grow faster, as emphasized by Demsetz (1973) and others. We present a structural model of imperfect competition emphasizing the relationship between size and efficiency. This structural model recognizes that lasting differences in firm size are caused by the fixity of capital, but differences in efficiency also matters. The model suggests how we can summarize and interpret different observable indicators of firm performance such as size, capital accumulation, productivity and profitability by a dynamic, latent factor which we call efficiency.

We use the term efficiency rather than productivity, as our structural model suggests that our measure of efficiency is unrelated to labor productivity. The argument is simple. Consider firms in a competitive industry with different levels of efficiency[1]. A firm with high efficiency will choose a high level of factor input so that its marginal product is the same as for the other 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). We show in section 4 that similar arguments hold beyond the special case with a Cobb-Douglas production function and price taking behavior, and the argument rises the question of how we should make inferences about differences

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

in efficiency from firm level data[2].

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 stochastic trends, initial conditions, transient noise, and industry-wide fluctuations. The multivariate framework imposes relatively few restrictions on the data generating process and allows us to consider the validity of the restrictions imposed by our structural model. According to the time-series terminology; our structural model of firm behavior requires that supply and factor inputs must 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 the fact that we do not explicitly model the firms' exit decisions.

2 A first look at patterns of firm performance

Are the 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, which are based on a rather complete, unbalanced sample of firm level observations from six (two-digit NACE) high-tech manufacturing industries. Details of the data are presented in Section 5. Figure 1 presents the means and standard deviations of log sales as a function of firm age[3]. All observations are measured relative to the industry-year mean. Not surprisingly, the graph shows that young firms are substantially smaller than older firms and that firm growth on average 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. It declines steadily to 0.76 when we correlate log sales in the first year and the 12'th year.

These patterns suggest that the differences across young firms are as large as among older firms and that they are highly persistent, emphasizing the role of intrinsic 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.

Another question we discussed in the introduction is how to measure firm performance? Figure 3 presents means and standard deviations of (log) labor productivity as a function of firm age and the

[2] A more extensive discussion of this issue is presented in Klette (2001). See also Bernard et al. (2000) and Klette and Kortum (2001).

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

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 differences between sales and labor productivity as measures of firm performance 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. The low correlation coefficient between productivity in the two first years suggests that almost half of the observed variance in labor productivity is temporary fluctuations or noisy data. A low degree of persistence in differences in labor productivity is not restricted to the firm's 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 at best is a rather noisy measure of efficiency.

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, but 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 Decomposing differences in firm performance

This section presents our econometric framework for assessing 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. We start by illustrating our approach in a simplified case before we present our complete multivariate framework.

3.1 A simplified, univariate framework

Stochastic specification: In the simplified version of our empirical framework, we focus on a univariate measure of firm performance; y_{it}. This could for instance be firm size, profits or some measure of productivity. Furthermore, we assume that all firms are established in year $t = 1$, and that the model have the following structure:

$$y_{it} = v_i + a_{it} + d_t + e_{it}, \quad t = 1, ..., T, \tag{1}$$

where

$$a_{it} = \begin{cases} 0 & t = 1 \\ a_{i,t-1} + \eta_{it} & t = 2, .., T, \end{cases} \tag{2}$$

and

$$v_i \sim \mathcal{IN}(0, \sigma_v^2), \quad \eta_{it} \sim \mathcal{IN}(0, \sigma_\eta^2), \quad e_{it} \sim \mathcal{IN}(0, \sigma_e^2). \tag{3}$$

The first two terms on the right hand side of (1) can be interpreted in terms of the two classes of models discussed in section 2: The term v_i is a random effect corresponding to the intrinsic differences in performance, while a_{it} is an idiosyncratic stochastic trend with innovations η_{it}. The last two terms in (1) represent sector-wide changes in performance, d_t, and (white noise) temporary fluctuations, e_{it}.

Note that a_{it} can be expressed as: $a_{it} = \sum_{s=2}^{t} \eta_{is}$, highlighting that it is cumulated sum of idiosyncratic innovations that are uncorrelated across firms and over time. In this simple model the assumption that $E\{\eta_{it}\} = 0$ is not restrictive, as any non-zero innovation mean will be absorbed into d_t. Since v_i has mean zero, and $a_{i1} = 0$, v_i is the time-invariant (intrinsic) part of the deviation of y_{it} from the industry mean d_t. Our analysis of why firms differ focuses on the relative importance of the first two terms on the right hand side of (1) in accounting for observed firm heterogeneity.

Measuring the importance of intrinsic differences and idiosyncratic innovations: To identify the sources of heterogeneity between firms, we decompose the variance in performance in year T . From equation (1)

$$\text{Var}\,(y_{iT}) = \sigma_v^2 + (T-1)\sigma_\eta^2 + \sigma_e^2.$$

¿From this expression, a natural measure of the importance of the intrinsic differences relative to the idiosyncratic innovations in explaining firm heterogeneity in year T is:

$$\frac{\text{Var}(v_i)}{\text{Var}(v_i + a_{iT})} = \frac{\sigma_v^2}{\sigma_v^2 + (T-1)\sigma_\eta^2}. \tag{4}$$

The ratio in (4) measures intrinsic differences as a fraction of the total variance in the non-transient performance in a given year. A closely related measure, which we label as the (unconditional) variance ratio, V, is:

$$V \equiv \frac{\text{Var}(v_i)}{\text{Var}(a_{iT})} = \frac{\sigma_v^2}{(T-1)\sigma_\eta^2} \tag{5}$$

– i.e. the ratio between the variance in intrinsic performance and the variance in cumulated innovations.

It is tempting to pursue the idea of variance decomposition – as in (4) – to quantify the relative importance of cumulated innovations and intrinsic heterogeneity. However, the idea of variance decomposition is not useful when v_i and a_{it} are correlated. Even if v_i and a_{it} are uncorrelated in our basic model (1), 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 (5): Let M_T be the set of firms that operate in year T. We define the *conditional variance* ratio, CV, as

$$CV \equiv \frac{\text{Var}\,(v_i | i \in M_T)}{\text{Var}\,(a_{it} |\ i \in M_T)}. \tag{6}$$

CV takes the endogeneity of firm exit into account. The computation of this measure in non-trivial as neither v_i or a_{it} are observable. We will return to the computational issues in Section 6, where we also elaborate our discussion of the self-selection problem and other econometric issues.

Preliminary remarks on identification: ¿From the covariance-matrix for y_{it} all parameters are identified from the cross sectional variation in the data, provided the sample covers at least two time periods:

$$\operatorname{Cov}(y_{it}, y_{is}) = \begin{cases} \sigma_v^2 + \sigma_\eta^2 \left[\min(t,s) - 1\right] & t \neq s \\ \sigma_v^2 + \sigma_\eta^2 (t-1) + \sigma_e^2 & t = s \end{cases} \tag{7}$$

As will be explained below, $\sigma_v^2, \sigma_\eta^2$ and σ_e^2 are three parameters of primary interest, and (7) shows that these parameters are identified from the covariance matrix for y_{it} with a sample covering at least two years.

Although identification of the model (1)-(3) appears almost trivial, the situation is complicated by the fact that we do not observe the firm over a fixed time period $[1, T]$, but from $t = \tau_i$ until $t = T_i$, where $\tau_i \geq 1$ is firm i's birth date, and $T_i \leq T$ is the last observation year: the exit time of firm i or the end of the survey period, T (whichever comes first). The stopping time T_i is an endogenous stochastic variable when survival is related to performance, y_{it}. This means that the operating firms are self-selected into the sample at any fixed t, which complicates identification and estimation as discussed in section 6.1.

3.2 Our multivariate econometric model

The multivariate model we estimate generalizes (1)-(3) by allowing the observable and unobservable variables to be vectors and it also introduces capital accumulation. Let $\mathbf{y}_{it} = (s_{it}, m_{it}, l_{it}, k_{it})'$, where s_{it}, m_{it}, and l_{it}, are (log) sales, material inputs, and labor, respectively, and k_{it} is log of the capital stock at the *end* of year t. The econometric model is:

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

where

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

and $\mathbf{v}_i, \boldsymbol{\eta}_{it}$ and $\mathbf{e}_{it}$ have independent, multivariate normal distributions:

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

Firm i is observed from year $\tau_i \geq 1$ until $T_i \leq T$, where the birth dates τ_i have an exogenous distribution, while the exit dates T_i may be endogenous (as specified in section 6.1).

In (8), $\mathbf{v}_i$ and $\mathbf{a}_{it}$ are 4-dimensional vectors of latent variables; $\boldsymbol{\gamma}_k$ is a 4×1 vector of capital coefficients corresponding to $k_{i,t-1}$ (the capital stock at the beginning of year t); and $\mathbf{e}_{it}$ is a 4×1 vector of white noise errors. Notice that from (9)-(10), we obtain $Var(\mathbf{a}_{it}) = (t - \tau_i)\Sigma_\eta$.

Since firms have varying birth dates, τ_i, any non-zero – and possibly non-constant – innovation mean will, in principle, be identifiable from the data. Thus; although we have assumed *a priori* that $\boldsymbol{\eta}_{it}$ are i.i.d. with zero mean throughout the life-time of a firm, these constraints are not necessarily satisfied in the (cross section of) *conditional* distributions of η_{it} (and a_{it}), given the observed variables on firm

i. The conditional distributions of the latent variables, which in fact are posterior distributions, to use Bayesian terminology, are the distributions of main interest in the empirical part of this paper. It is well known from Bayesian statistics that inferences based on posterior distributions are generally robust with respect to moderate alternations of the prior distribution. Our analysis is in line with the empirical Bayes approach to modelling stochastic trends, which is well established in the time series literature (see for example Kitagawa (1996)).

An important aspect of the multivariate model is to account for the role of *capital accumulation* in the process of firm growth. As we shall see in section 4, a restricted version of the multivariate model can be given a structural interpretation in terms of an explicit model of firm behavior. According to the analysis in section 4, the capital variable plays a crucial role by allowing us to interpret the $\mathbf{v}_i$ and $\mathbf{a}_{it}$ as differences in efficiency.

Variance decomposition in the multivariate case: We want to interpret the cross-sectional heterogeneity of firms in year T, similar to our analysis in section 3.1 . This can be done by multivariate variance decomposition using the trace of the variance-covariance matrices as a measure of multivariate variance. In the general rank-r case, we obtain

$$\text{tr } Var\{\mathbf{y}_{iT}\} = \text{tr } \Sigma_v + (T - \tau_i) \text{ tr } \Sigma_\eta + \text{tr } \Sigma_e,$$

Let $\bar{T}$ denote the average age of firms operating in year T: $\bar{T} = E\{T - \tau_i | i \in M_T\}$. Hence, the natural extensions of our univariate measures of the importance of intrinsic differences relative to idiosyncratic innovations are:

$$V = \frac{\text{tr } \Sigma_v}{\bar{T} \text{ tr } \Sigma_\eta}$$

and

$$CV = \frac{\text{tr Var}(\mathbf{v}_i | i \in M_T)}{\text{tr Var}(\mathbf{a}_{iT} | i \in M_T)}.$$

One or more latent components? In the multivariate model, an interesting question is: What is the number of independent stochastic components in $\mathbf{v}_i$ and $\mathbf{a}_{it}$, respectively? The answer is determined by the rank of the covariance matrices Σ_v and Σ_η. For example, if the rank of Σ_η is 1, 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 vector of fixed coefficients. Thus, $\mathbf{a}_{it} = \mathbf{u}_\eta a_{it}$, where $a_{it} = \sum_{s \le t} \eta_{is}$. In fact, our empirical results show that a one-factor model accounts for more than 90% of the variation in $\mathbf{a}_{it}$ in all the industries we consider.

In the general case, the rank of Σ_η is r ($r \le 4$), and we can represent the innovations η_{it} through an orthogonal factor decomposition (see Anderson, 1984):

$$\boldsymbol{\eta}_{it} = \mathbf{u}_{\eta,(1)} \eta_{it,(1)} + ... + \mathbf{u}_{\eta,(r)} \eta_{it,(r)}, \tag{11}$$

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 $\sigma^2_{\eta,(1)} \geq ... \geq \sigma^2_{\eta,(r)} > 0$. In (11), the $\eta_{it,(j)}$ are latent variables with:

$$\begin{aligned} E\{\eta_{it,(j)}\} &= 0 \\ Cov\{\eta_{it,(j)}, \eta_{it,(k)}\} &= \begin{cases} \sigma^2_{\eta,(j)} & j = k \\ 0 & j \neq k. \end{cases} \end{aligned} \tag{12}$$

Hence, $\Sigma_\eta = \sum_{j=1}^{r} \mathbf{u}_{\eta,(j)} \mathbf{u}_{\eta,(j)}{}' \sigma^2_{\eta,(j)}$. We will refer to the vector $\mathbf{u}_{\eta,(j)}\sigma_{\eta,(j)}$ as the *loadings* on the j'th factor of Σ_η. Note that

$$\mathbf{a}_{it} = \mathbf{u}_{\eta,(1)} a_{it,(1)} + ... + \mathbf{u}_{\eta,(r)} a_{it,(r)},$$

where $a_{it,(j)} = \sum_{s \leq t} \eta_{is,(j)}$ for $j = 1, .., r$.

A similar analysis can be done with respect to $\mathbf{v}_i$:

$$\mathbf{v}_i = \mathbf{u}_{v,(1)} v_{i,(1)} + ... + \mathbf{u}_{v,(r)} v_{i,(r)},$$

where

$$\begin{aligned} E\{v_{i,(j)}\} &= 0 \\ Cov\{v_{i,(j)}, v_{i,(k)}\} &= \begin{cases} \sigma^2_{v,(j)} & j = k \\ 0 & j \neq k \end{cases} \end{aligned} \tag{13}$$

Consequently, the loadings on the j'th factor of Σ_v is given by the vector $\mathbf{u}_{v,(j)}\sigma_{v,(j)}$. In the next section a structural interpretation of this model will be presented which puts restrictions on the rank r of Σ_η and Σ_v, and on the corresponding factor loadings.

4 A structural interpretation of our multivariate model

The multivariate framework presented in section 3.2 can be given a structural interpretation, based on a model of firm behavior expressing supply and factor demand as explicit functions of efficiency, demand, capital, and factor prices. This section shows how efficiency and demand differences across firms affect supply and factor demand, imposing strong restrictions on the parameters in the multivariate model presented in section 3.2.

Supply and factor demand: Consider the production function

$$Q_{it} = A_{it} K^{\gamma}_{i,t-1} F\left(M_{it}, L_{it}\right), \tag{14}$$

where Q_{it}, A_{it} and $K_{i,t-1}$ denote the firm's output, efficiency and capital, while $F(M_{it}, L_{it})$ is a function aggregating materials and labor inputs. $F(M_{it}, L_{it})$ is homogenous of degree ε. Given factor prices for labor and materials common across firms, w^l_t and w^m_t, and treating $K_{i,t-1}$ as pre-determined, it follows that the short-run cost-function has the following form:

$$C(w^m_t, w^l_t, Q_{it}, A_{it}, K_{i,t-1}) = G(w^m_t, w^l_t) \left(\frac{Q_{it}}{A_{it} K^{\gamma}_{i,t-1}} \right)^{1/\varepsilon}. \tag{15}$$

The factor demand for materials and labor is then

$$\ln M_{it} = \ln \frac{\partial C}{\partial w_t^m} = \frac{1}{\varepsilon}\left(\ln Q_{it} - \ln A_{it}\right) - \frac{\gamma}{\varepsilon} \ln K_{i,t-1} + \ln G_{t,m} \tag{16}$$

$$\ln L_{it} = \ln \frac{\partial C}{\partial w_t^m} = \frac{1}{\varepsilon}\left(\ln Q_{it} - \ln A_{it}\right) - \frac{\gamma}{\varepsilon} \ln K_{i,t-1} + \ln G_{t,l} \tag{17}$$

where $G_{t,m} = \partial G(w_t^m, w_t^l)/\partial w_t^m$ and $G_{t,l} = \partial G(w_t^m, w_t^l)/\partial w_t^l$. Each firm faces a demand function

$$Q_{it} = D_{it} P_{it}^{-e}, \tag{18}$$

where D_{it} denotes the level of demand, and each firm charges a price which is a markup, μ_{it} ($\mu_{it} \geq 1$), times marginal costs:

$$P_{it} = \mu_{it} \frac{\partial C}{\partial Q_{it}}. \tag{19}$$

We now want to derive factor demand and supply as a function of the capital, technological level and the markup. We do not observe supply in terms of output, but instead in terms of sales, $S_{it} = P_{it} Q_{it}$. Combining this expression with (15), (16), (17), (18) and (19), the set of supply and (short-run) factor demand equations can then be stated

$$\begin{bmatrix} \ln S_{it} \\ \ln M_{it} \\ \ln L_{it} \end{bmatrix} = \frac{e-1}{\varepsilon + e - e\varepsilon} \left\{ \mathbf{1} \ \ln A_{it}^* + \mathbf{1}\ \gamma \ \ln K_{i,t-1} - \begin{bmatrix} \varepsilon \\ (1-1/e)^{-1} \\ (1-1/e)^{-1} \end{bmatrix} \ln \mu_{it} \right\} + \mathbf{g}(w_t^m, w_t^l) \tag{20}$$

where $\mathbf{1} = [1,1,1]'$ and $A_{it}^* = D_{it}^{1/(e-1)} A_{it}$. The last term on the right hand side is a vector common across firms which may vary over time as it depends (only) on the common factor prices. A_{it} and D_{it} enter this system of equations symmetrically, and we are consequently not be able to distinguish between technology and demand shocks in the empirical analysis where they both are unobservable[4].

We will refer to A_{it}^* as a firm's *efficiency*, and this index embodies both product and process innovations cumulated from the firm's past. (20) suggests that differences in labor productivity, i.e. value added per labor input $(S_{it} - M_{it})/L_{it}$, are independent of differences in firm efficiency, A_{it}^* . We notice, however, that firm level differences in markups will show up as differences in labor productivity[5].

According to this model, differences in firm size, conditional on their different capital stocks, are informative about differences in firm efficiency[6]. This relationship between 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. Furthermore, it can at least partly explain why average size trends upward among the surviving firms, while no similar trend is visible in labor productivity, as shown in Figures 1 and 3.

[4] This is also pointed out by Levinsohn and Melitz (2001).

[5] That is, $\ln\left[(S_{it} - M_{it})/L_{it}\right] = \ln \mu_{it}$ ignoring terms related to $\mathbf{g}(w_t^m, w_t^l)$. We notice that differences in labor productivity and efficiency may be correlated if differences in efficiency are correlated with differences in markups.

[6] This is even more transparent in with the case perfect competition, i.e. with $\mu_{it} = 1$ and $e \to \infty$, where (20) simplifies to

$$\begin{bmatrix} \ln S_{it} \\ \ln M_{it} \\ \ln L_{it} \end{bmatrix} = \frac{1}{1-\varepsilon}\left(\mathbf{1} \ \ln A_{it} + \gamma \mathbf{1} \ \ln K_{it}\right).$$

ignoring the common term $\lim_{e\to\infty} \mathbf{g}(w_t^m, w_t^l)$.

Monopolistic competition In the rest of this section, we will focus on industries with monopolistic competition, where $\mu_{it} = (1 - 1/e)^{-1}$ for all firms i. In this case, the last two terms on the right hand side of (20) do not vary across firms, and short-run profits can be expressed in terms of A^*_{it} and $K_{i,t-1}$

$$\Pi_t(A^*_{it}, K_{i,t-1}) = S_{it} - M_{it} - L_{it} = \pi_t \; [A^*_{it} K_{i,t-1}]^{(e-1)/(\varepsilon+e-e\varepsilon)},$$

where π_t is a factor common across firms (which may vary across years and industries). According to this expression there is an increasing, one-to-one relationship between the differences in profits not accounted for by capital, and the differences in efficiency.

Capital stock dynamics: Consider now the capital stock dynamics derived from each firm's investment behavior. Let $I(K_{it}, K_{i,t-1})$ denote the costs of changing 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. This investment function is convex in K_{t-1} and decreasing in K_t. In Appendix C, we examine the firm's investment problem, which can be formulated in terms of dynamic programming

$$V(A^*_{it}, K_{i,t-1}) = \max_{K_{it}} \left\{ \Pi_t(A^*_{it}, K_{i,t-1}) - I(K_{it}, K_{i,t-1}) + \beta E \left[V(A^*_{i,t+1}, K_{it}) | A^*_{it}, K_{i,t-1} \right] \right\} \tag{21}$$

where $V(A^*_{it}, K_{i,t-1})$ is the value function and $E[\cdot | A^*_{it}, K_{i,t-1}]$ is the expectation conditional on A^*_{it} and $K_{i,t-1}$. Appendix C shows that a log-linear approximation to the policy function corresponding to (21), is

$$\ln K_{it} = \alpha_t + \alpha_a \ln A^*_{i,t-1} + \alpha_k \ln K_{i,t-1}. \tag{22}$$

where α_a is a positive parameter, α_k is a parameter between zero and one, and α_t captures changes in factor prices and other industry-wide variables alternating over time.

The structural model of firm growth: Combining (20) and (22)

$$\mathbf{y}_{it} = \boldsymbol{\theta}_a \; \ln A^*_{i1} + \boldsymbol{\theta}_a \; \ln (A^*_{it}/A^*_{i1}) + \boldsymbol{\theta}_k \ln (K_{i,t-1}) + \boldsymbol{\theta}_t. \tag{23}$$

where

$$\begin{aligned} \mathbf{y}_{it} &\equiv \begin{bmatrix} \ln S_{it} & \ln M_{it} & \ln L_{it} & \ln K_{it1} \end{bmatrix}' \\ \boldsymbol{\theta}_a &= \begin{bmatrix} \frac{(e-1)}{\varepsilon+e-e\varepsilon}, & \frac{(e-1)}{\varepsilon+e-e\varepsilon}, & \frac{(e-1)}{\varepsilon+e-e\varepsilon}, & \alpha_a \end{bmatrix}' \\ \boldsymbol{\theta}_k &= \begin{bmatrix} \frac{\gamma(e-1)}{\varepsilon+e-e\varepsilon}, & \frac{\gamma(e-1)}{\varepsilon+e-e\varepsilon}, & \frac{\gamma(e-1)}{\varepsilon+e-e\varepsilon}, & \alpha_k \end{bmatrix}' \end{aligned} \tag{24}$$

while $\boldsymbol{\theta}_t = \begin{bmatrix} \mathbf{g}'(w^m_t, w^l_t), & \alpha_t \end{bmatrix}'$.

The model (23)-(24) suggests that the *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*, $\boldsymbol{\theta}_k \ln (K_{i,t-1})$. Capital accumulation is, according to (23), driven by cumulated changes in efficiency and independent stochastic shocks. In (23) differences in efficiency are decomposed into two components: intrinsic differences introduced already when the firms are born, $\boldsymbol{\theta}_a \ln A^*_{i1}$, and differences in the cumulated changes in (relative) efficiency, $\boldsymbol{\theta}_a \ln (A^*_{it}/A^*_{i1})$.

The model of firm behavior presented in this section is highly constrained as it assumes that efficiency changes affect all the components of $\mathbf{y}_{it}$ through a single factor. And not only is the model limited to one latent factor, – the three first loading-components on this latent factor are equal, which is a strong restriction on the multivariate model presented in section 3.2. That is, in the multivariate, descriptive model there are no a priori constraints on the covariance matrices Σ_v and Σ_η (apart from positive semi-definiteness), while, according to the structural model presented in this section, these two matrices are rank 1 and can be factorized as:

$$\begin{aligned} \Sigma_v &= \boldsymbol{\theta}_a \boldsymbol{\theta}_a' \, Var\left(\ln A_{i1}^*\right) \\ \Sigma_\eta &= \boldsymbol{\theta}_a \boldsymbol{\theta}_a' \, Var\left[\ln\left(A_{it}^*/A_{i1}^*\right)\right]. \end{aligned} \tag{25}$$

In the time series terminology, our structural model imposes a co-integration relationship between the three first components of $\mathbf{y}_{it}$, with a highly constrained co-integration vector. Below we will examine the empirical validity of these constraints. Notice that $\boldsymbol{\theta}_a Var\left(\ln A_{i1}^*\right)^{\frac{1}{2}}$ and $\boldsymbol{\theta}_a Var\left[\ln\left(A_{it}^*/A_{i1}^*\right)\right]^{\frac{1}{2}}$ are the factor loadings in the rank-1 decomposition of Σ_v and Σ_η, respectively.

¿From equations (23)-(24), we see that none of the structural parameters in the production or the demand function is identified without additional restrictions. Nevertheless, the structural model imposes a lot of constraints on the covariance matrix and auto-correlation structure in the observed data[7]. This is true even if we assume perfect competition.

Industry wide changes in efficiency and demand are absorbed in the $\boldsymbol{\theta}_t$-term. Transient deviations from this model of firm behavior and noise in the data correspond to the $\mathbf{e}_{it}$-term in the multivariate model in section 3.2 .

To summarize, this section has shown that the multivariate model presented in section 3.2 has a structural interpretation when the number of latent factors are reduced to one and the factor loadings satisfy the constraints presented above. According to this model, differences in observed performance, $\mathbf{y}_{it}$, are driven by differences in efficiency across firms and by differences in their capital stocks. The econometric framework presented in section 3.2 extracts transient shocks and industry wide changes in firms' performance, and allow us to focus on the evolution of the persistent differences in efficiency across firms. The structural model decomposes these differences in efficiency into intrinsic differences and differences which are cumulated as the firms evolve. Differences in capital accumulation are at least partly driven by efficiency differences across firms.

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.

[7] Notice that $\boldsymbol{\theta}_a$, and consequently γ, is not identified, because we do not observe the (variance of) $\ln A_{it}^*$.

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 establishments 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[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 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 which were operating during 1973-95 were included in the sample, and observed until $T = 1995$. However, the sample of interest consist only of the firms born within the 24 year period. For the other firms separate (nuisance) parameters were estimated for the distribution of v_i[10]. The reason is that for these firms, v_i is composed of both intrinsic differences and cumulated innovations (up until 1973) and therfore has a different meaning than for firms in the interest sample, i.e. it is not a time-invariant variable. For this reason, when we later analyse firm heterogeneity, establishments entering the industry before 1973 were excluded from the analysis. Of *all* plants operating in 1995, 75-85 percent were established after 1972, and thus included in the interest sample. These firms account for a similar share of total sales in 1995 (measured on log-scale for each firm).

Some "cleaning" of the data was also performed: A firm was excluded from the sample if either; (i) the value of an endogenous variable is missing for two subsequent years or more; (ii) the firm disappears from the raw data file and then reappears; or (iii) the firm is observed for less than 2 years. These trimming procedures reduced the data set by 15-20 percent. In addition we removed firms with extreme variation[11] in the endogenous variables, which eliminated an additional 4-8 percent of the observations.

Some summary statistics are presented in Table 1. Table 1 reveals considerable heterogeneity within industries in terms of the variation in sales. The size distribution is highly skewed within each sector; mean-sales is much higher than median-sales.

[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 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.

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

[11] 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.

6 Econometric issues

Our econometric model, presented in section 3, raises a set of econometric issues which we address in this section. These issues are: (i) identification in the presence of self-selection when we do not explicitly model the exit-process, (ii) estimation of the structural parameters of the model, and (iii) calculation of the conditional variance ratio 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 Selection and identification

We presented a preliminary discussion of how the model is identified in section 3.1 focusing on a uni-variate version of the model. In this section, we return to a discussion of identification issues in relation to the uni-variate model (1)-(3). The generalization to the multivariate raises no new issues.

The results of Cox ? and Little and Rubin (1987), shown that a partial likelihood – that is, the (pseudo) likelihood function obtained by treating the stopping times T_i as if they were fixed or independent og y_{it} when setting up the likelihood function – provide consistent estimators in the presence of systematic selection, if the stochastic process, y_{it}, satisfies the so-called missing at random (MAR) condition[12]. The version of the MAR condition needed in our case is:

$$f(y_{i,t+1}|\mathrm{I}(t < T_i), y_{i1}, .., y_{it}; \boldsymbol{\beta}) = f(y_{i,t+1}|y_{i1}, .., y_{it}; \boldsymbol{\beta}),\ t = 1, ..., T \text{ and } i = 1, .., N, \tag{26}$$

where $f(\cdot|\cdot)$ is generic notation for conditional probability density, $\mathrm{I}(t < T_i)$ is the indicator variable which is 1 if the firm will operate also in year $t + 1$, and 0 otherwise, and $\boldsymbol{\beta}$ is the model parameters. Equation (26) says that information about *survival* up until time t should not help prediction of $y_{i,t+1}$ given $y_{i1}, ..., y_{it}$. The important property is that the observation of $\mathrm{I}(t < T_i)$ cannot be used to improve predictions about next years' performance, $y_{i,t+1}$, given the history of the y_{it}-process up until t [13]. A situation where MAR fails is, say, if the firm knows at the end of year t what the sales will be in year $t + 1$, and chooses to exit if the (potential) sales is below some threshold. In this case, the value of $\mathrm{I}(t < T_i)$ gives information about $y_{i,t+1}$ not being contained in $y_{i1}, ..., y_{it}$. The MAR assumption is the most general self-selection mechanism which allows identification based on the partial likelihood.

Identification of $\boldsymbol{\beta}$ based on the partial likelihood function are achieved provided (26) holds and $\boldsymbol{\beta}$ is identified in the model without attrition. This result holds even if exit depends on β, as discussed in Raknerud (2001a). We will hereafter use the term likelihood when, in fact, we consider a partial (or 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)

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

[13] When a firm exits in year $t + 1$, s_{it+1}, denotes the sales the firm could have obtained in that year if it hadn't exited. Notice that the MAR assumption does not exclude firms from having private information which affect their exit decisions, e.g. information about scrap values.

method of moments[14].

6.2 Estimation

The main tool in the estimation of the parameters is a generalization of the EM (Expectation Maximization) algorithm called the ECM (Expectation Conditional Maximization) algorithm (Meng, 1993). The parameters, $\boldsymbol{\beta}$, to be estimated are $(\Sigma_\eta, \Sigma_v, \Sigma_e, \boldsymbol{\gamma}_k, \mathbf{d})$, where $\mathbf{d}$ denotes the time dummies $(\mathbf{d}_1, ..., \mathbf{d}_T)$. The EM algorithm was originally developed by Dempster et al. (1977) as a tool for estimating models with incomplete or missing data when the likelihood of the complete data has a simple explicit form. In our case, $\mathbf{y}_{it}$ is the observed (incomplete) data, while $(\mathbf{v}_i{}', \mathbf{a}_{i,\tau_i}{}', ..., \mathbf{a}_{i,T_i}{}')$ is missing, and the likelihood of the complete data is determined by the distribution of $\sum_{i=1}^N (T_i - \tau_i + 1)$ i.i.d. terms $\mathbf{e}_{it}$.

The state space representation: Using a Helmert-type orthogonal transformation of the variables, we show that $\mathbf{v}_i$ can be "integrated out" of the likelihood function, as originally proposed in Raknerud (2001b). This transformation substantially reduces the number of latent variables, which is important to obtain rapid convergence of the estimation algorithm. Moreover, the resulting concentrated likelihood can be expressed on a state space form, where the state vector is $\boldsymbol{\alpha}_{is} = (\mathbf{a}_{i,t+1}{}', \sum_{s=\tau_i}^t \mathbf{a}_{i,s}{}')'$ for $t = \tau_i, ..., T_i$.

To obtain a state space form of the model (8)-(10) which is useful for estimation purposes, we start by defining

$$\mathbf{R}_{it} = \mathbf{v}_i + \mathbf{e}_{it}.$$

We use a tilda, e.g. $\widetilde{\mathbf{R}}_{it}$, to indicate the following Helmert-type transformation:

$$\widetilde{\mathbf{R}}_{is} = \begin{cases} \sqrt{\frac{s}{s+1}}(\mathbf{R}_{i,\tau_i+s} - \frac{1}{s}\sum_{v=0}^{s-1} \mathbf{R}_{i,\tau_i+v}) & s = 1, .., T_i - \tau_i \\ \frac{1}{s}\sum_{v=0}^{s-1} \mathbf{R}_{i,\tau_i+v} & s = T_i - \tau_i + 1. \end{cases} \tag{27}$$

Note that index s is the age of firm i at time t: $s = t - \tau_i + 1$. The transformation also applies to sequences of variables that are not firm-specific. For example

$$\widetilde{\mathbf{d}}_{is} = \begin{cases} \sqrt{\frac{s}{s+1}}(\mathbf{d}_{\tau_i+s} - \frac{1}{s}\sum_{v=0}^{s-1} \mathbf{d}_{\tau_i+v}) & s = 1, .., T_i - \tau_i \\ \frac{1}{s}\sum_{v=0}^{s-1} \mathbf{d}_{\tau_i+v} & s = T_i - \tau_i + 1. \end{cases}$$

It is easy to verify that

$$\begin{aligned} E\{\widetilde{\mathbf{R}}_{is}\} &= \mathbf{0}, \quad s = 1, .., T_i - \tau_i + 1 \\ Var\{\widetilde{\mathbf{R}}_{is}\} &\equiv \begin{cases} \Sigma_e & s = 1, .., T_i - \tau_i \\ \Sigma_v + s^{-1}\Sigma_e & s = T_i - \tau_i + 1 \end{cases} \\ E\{\widetilde{\mathbf{R}}_{is}\widetilde{\mathbf{R}}_{it}{}'\} &= \mathbf{0}, \quad s \neq t. \end{aligned} \tag{28}$$

The state space form of the model with state vector $\boldsymbol{\alpha}_{is}$ becomes (see Raknerud (2001b) for a more detailed account):

$$\begin{aligned} &\widetilde{\mathbf{y}}_{is} = \mathbf{G}_{is}\boldsymbol{\alpha}_{is} + \widetilde{\mathbf{d}}_{is} + \boldsymbol{\gamma}_k \widetilde{k}_{is} + \widetilde{\mathbf{R}}_{is} \\ &\boldsymbol{\alpha}_{is} = F\boldsymbol{\alpha}_{i,s-1} + \boldsymbol{\omega}_{is} \end{aligned} \qquad s = 1, ..., T_i - \tau_i + 1, \tag{29}$$

[14] The reason for this is that the covariance structure (7) cannot be estimated from sample analogues: If exit is endogenous, $\text{Cov}(s_{it}, s_{is}| \max(s,t) \leq T_i)$ will not in general be given by (7) (even if MAR holds). Hence the sample covariance matrix ceases to provide consistent estimators for the model parameters.

where

$$
\begin{aligned}
\mathbf{G}_{is} &= \begin{cases} \left[\begin{array}{cc} \sqrt{\frac{s}{s+1}}\mathbf{I}_r & -\sqrt{\frac{1}{s(s+1)}}\mathbf{I}_r \end{array} \right] & s = 1, ..., T_i - \tau_i \\ \left[\begin{array}{cc} \mathbf{0} & \frac{1}{s}\mathbf{I}_r \end{array} \right] & s = T_i - \tau_i + 1 \end{cases} \\
\boldsymbol{\alpha}_{i0} &= \mathbf{0} \\
F &= \left[\begin{array}{cc} \mathbf{I}_r & \mathbf{0} \\ \mathbf{I}_r & \mathbf{I}_r \end{array} \right] \\
\boldsymbol{\omega}_{is} &= \left[\begin{array}{c} \boldsymbol{\eta}_{i,\tau_i+s} \\ \mathbf{0} \end{array} \right], \quad s = 1, ..., T_i - \tau_i + 1.
\end{aligned}
$$

($\mathbf{I}_r$ is the identity matrix of order r). Note that the $\boldsymbol{\omega}_{is}$, from (10), are uncorrelated with mean zero and singular covariance matrix:

$$
E\{\boldsymbol{\omega}_{is}\boldsymbol{\omega}_{is}'\} \equiv \left[\begin{array}{cc} \Sigma_\eta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array} \right], \; s = 1, ..., T_i - \tau_i + 1. \tag{30}
$$

and that $\boldsymbol{\omega}_{is}$ and $\widetilde{\mathbf{R}}_{it}$ are independently distributed for all s and t.

The outline of the ECM algorithm: Given the state space representation (29), it is straightforward to apply the ECM algorithm to maximize the likelihood function. We here present a general outline of the algorithm, while computational details are deferred to Appendix E.

For random vectors $\mathbf{y}$ and $\boldsymbol{\alpha}$, let $\mathbf{y}$ denote the observed ("incomplete") data and $\boldsymbol{\alpha}$ the "missing" data. Furthermore, let $g(\mathbf{y}, \boldsymbol{\alpha}; \boldsymbol{\beta})$ be their joint density (i.e. the "complete" data density), and $g(\boldsymbol{\alpha}|\, \mathbf{y}; \boldsymbol{\beta})$ the conditional density of $\boldsymbol{\alpha}$ given $\mathbf{y}$. The ML estimator, $\widehat{\boldsymbol{\beta}}$, is the maximum of the log-likelihood $L(\boldsymbol{\beta})$ of the observed data, where

$$
L(\boldsymbol{\beta}) = \ln g(\mathbf{y}; \boldsymbol{\beta}). \tag{31}
$$

Since

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

(31) can be rewritten as

$$
L(\boldsymbol{\beta}) = \ln g(\mathbf{y}, \boldsymbol{\alpha}; \boldsymbol{\beta}) - \ln g(\boldsymbol{\alpha}|\, \mathbf{y}; \boldsymbol{\beta}). \tag{32}
$$

Taking the expectation on both sides in (32) with respect to $g(\boldsymbol{\alpha}|\, \mathbf{y}; \boldsymbol{\beta}')$, where $\boldsymbol{\beta}'$ is an arbitrary parameter value, gives:

$$
L(\boldsymbol{\beta}) = M(\boldsymbol{\beta}|\boldsymbol{\beta}') - H(\boldsymbol{\beta}|\, \boldsymbol{\beta}'), \tag{33}
$$

where

$$
\begin{aligned}
M(\boldsymbol{\beta}|\boldsymbol{\beta}') &= \int \ln g(\mathbf{y}, \boldsymbol{\alpha}; \boldsymbol{\beta})\, g(\boldsymbol{\alpha}|\, \mathbf{y}; \boldsymbol{\beta}')\, d\boldsymbol{\alpha} \\
H(\boldsymbol{\beta}|\boldsymbol{\beta}') &= \int \ln g(\boldsymbol{\alpha}|\mathbf{y}; \boldsymbol{\beta})\, g(\boldsymbol{\alpha}|\, \mathbf{y}; \boldsymbol{\beta}')\, d\boldsymbol{\alpha}.
\end{aligned}
$$

It is shown in Wu (1983) that the following EM algorithm will converge to a stationary point on the likelihood function under quite general conditions:

Let $\boldsymbol{\beta}^{(1)}$ be given. For $m = 1, 2, ...$

(i) E-step: Compute $M(\boldsymbol{\beta}|\, \boldsymbol{\beta}^{(m)})$.

(ii) M-step: Set $\boldsymbol{\beta}^{(m+1)} = \arg\max_{\boldsymbol{\beta}} M(\boldsymbol{\beta}|\, \boldsymbol{\beta}^{(m)})$.

(iii) Set $m = m + 1$, and go to *(i)*

Since by Kullback's inequality $H(\boldsymbol{\beta}|\boldsymbol{\beta}^{(m)}) \leq H(\boldsymbol{\beta}^{(m)}|\boldsymbol{\beta}^{(m)})$ for all $\boldsymbol{\beta}$, it is easy to verify that $\{L(\boldsymbol{\beta}^{(m)})\}$ is an increasing sequence of likelihood values.

It will be more convenient for us to replace the maximization in the M-step by a *conditional maximization* (CM) step: We partition the parameters in two blocks, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ as specified in Appendix E, and replace the M-step by two partial maximizations. That is, we first maximize $M(\boldsymbol{\beta}|\, \boldsymbol{\beta}^{(m)})$ with respect to $\boldsymbol{\beta}_1$ keeping $\boldsymbol{\beta}_2$ fixed at its current estimate. Then we maximize with respect to $\boldsymbol{\beta}_2$, with $\boldsymbol{\beta}_1$ fixed at its new, updated value. This partial maximization procedure guarantees that $M(\boldsymbol{\beta}^{(m+1)}|\boldsymbol{\beta}^{(m)}) > M(\boldsymbol{\beta}^{(m)}|\boldsymbol{\beta}^{(m)})$ and retains the main convergence properties of the EM algorithm (see Meng and Rubin (1993)).

The EM (ECM) algorithm does not require calculation of the log-likelihood $L(\boldsymbol{\beta})$ – only the function $M(\boldsymbol{\beta}|\boldsymbol{\beta}^{(m)})$. Another important property of the algorithm is that

$$\frac{\partial L(\boldsymbol{\beta}^{(m)})}{\partial \boldsymbol{\beta}} = \left.\frac{\partial M(\boldsymbol{\beta}|\boldsymbol{\beta}^{(m)})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(m)}}, \tag{34}$$

which follows from (33) and the fact that $\boldsymbol{\beta}^{(m)}$ is the maximizer of $H(\boldsymbol{\beta}|\boldsymbol{\beta}^{(m)})$, and hence a stationary point. The Hessian of $L(\boldsymbol{\beta})$ at $\widehat{\boldsymbol{\beta}}$ can therefore be obtained by numerical differentiation of $\left.\frac{\partial M(\boldsymbol{\beta}|\widehat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}}$. This result is important, because it yields a computationally simple estimator of the covariance matrix of $\widehat{\boldsymbol{\beta}}$.

6.3 Calculation of the conditional variance ratio

The conditional variance ratio (CV), defined in (6), is the ratio of the variances for the unobservables, i.e. $\text{Var}(\mathbf{v}_i|i \in M_T)$ and $\text{Var}(\mathbf{a}_{it}|i \in M_T)$, and must consequently be estimated, given estimates of the parameters $\boldsymbol{\beta}$. This section explains how $\text{Var}(\mathbf{v}_i|i \in M_T)$ and $\text{Var}(\mathbf{a}_{it}|i \in M_T)$ are estimated.

We see that equations (8)-(9) can readily be written on a state space form:

$$\begin{aligned}
\mathbf{y}_{it} &= \begin{bmatrix} \mathbf{I} & \mathbf{I} \end{bmatrix} \mathbf{x}_{it} + \boldsymbol{\gamma}_k k_{i,t-1} + \mathbf{d}_t + \mathbf{e}_{it},\ t = \tau_i, ..., T_i \\
\mathbf{x}_{it} &= \begin{cases} \begin{bmatrix} \mathbf{0} \\ \mathbf{v}_i \end{bmatrix} & t = \tau_i \\ \mathbf{x}_{i,t-1} + \begin{bmatrix} \boldsymbol{\eta}_{it} \\ \mathbf{0} \end{bmatrix} & t = \tau_i + 1, ..., T \end{cases}
\end{aligned}$$

where $\mathbf{x}_{it} = (\mathbf{a}_{it}{}', \mathbf{v}_i{}')'$, while $\mathbf{0}$ and $\mathbf{I}$ denote zero and identity matrices (of approperiate dimension), respectively. The Kalman filter and -smoother provide the mean, $\widehat{\mathbf{x}}_{iT|T}$, and covariance matrix, $\widehat{\mathbf{V}}_{iT|T}$, for the state vector $\mathbf{x}_{it}$ given $\mathbf{y}_{i,\to T} = (\mathbf{y}_{i,\tau_i}, ..., \mathbf{y}_{iT})$ (see e.g. Harvey (1989)):

$$\widehat{\mathbf{V}}_{iT|T} \equiv Var(\mathbf{x}_{iT}|\mathbf{y}_{i,\to T}) \text{ and } \widehat{\mathbf{x}}_{iT|T} \equiv E(\mathbf{x}_{iT}|\mathbf{y}_{i,\to T}).$$

However, to calculate CV we need an expression for

$$Var\left(\mathbf{B}\mathbf{x}_{iT}|i\in M_T\right)=\mathbf{B}Var\left(\mathbf{x}_{iT}|i\in M_T\right)\mathbf{B}'$$

for $\mathbf{B}=\left[\begin{array}{cc}\mathbf{I} & \mathbf{0}\end{array}\right]$ and $\mathbf{B}=\left[\begin{array}{cc}\mathbf{0} & \mathbf{I}\end{array}\right]$ (both $\mathbf{0}$ and $\mathbf{I}$ are of dimension 4×4).

It follows from the MAR assumption that:

$$f(\mathbf{x}_{iT}|i\in M_T,\mathbf{y}_{i,\to T})=f(\mathbf{x}_{iT}|\mathbf{y}_{i,\to T}). \tag{35}$$

By the rule of iterated expectation, and (35):

$$Var\{\mathbf{x}_{iT}|i\in M_T\}=E\{\widehat{\mathbf{V}}_{iT|T}|i\in M_T\}+Var\{\widehat{\mathbf{x}}_{iT|T}|i\in M_T\}.$$

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

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

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

7 Empirical results

This section presents our empirical results, which focus on the heterogeneity in performance among firms within six manufacturing industries in 1995. The results can be divided into two parts. First, we argue that our structural model presented in section 4 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 examine the differences in efficiency and how they evolve over time. In all the industries we consider, intrinsic differences are larger than the differences generated by cumulated, firm-specific innovations. Young firms are, on average, more innovative than older firms and they have a much larger variance in their innovations. Comparing the results across different industries, we find that the largest intrinsic differences are found in the industries that also have the largest differences in the idiosyncratic innovations. Finally, we examine how selection systematically eliminates firms with low efficiency.

7.1 A single factor model with a structural interpretation

The results in Table 2 and 3 support our simple, structural model presented in section 4, 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.

The last columns of Table 2 and 3 presents a pseudo R^2-measure:

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

where $\widehat{\mathbf{e}}_{it} = \mathbf{y}_{it} - E(\mathbf{v}_i + \mathbf{a}_{it}|$ all the data on firm $i) - \widehat{\gamma}_k k_{i,t-1} - \widehat{\mathbf{d}}_t$ (the expectation is evaluated at the estimated parameters and $\widehat{Var}(\cdot)$ denote the sample variance). R^2 varies between $.94 - .95$ in the one-factor model; and between $.97 - .98$ in the four factor model, underlining the excellent fit of the model with one latent factor.

These patterns of eigenvalues show that the persistent differences in performance can be summarized by the first latent factors $v_{i,(1)}$ and $a_{it,(1)}$, accounting for at least 90 percent of the variation in $\mathbf{v}_i$ and $\mathbf{a}_{it}$, respectively. We conclude that a single latent time-invariant component and a single latent random walk component, is largely adequate as a summary of firm performance[15].

Table 3 presents the estimated factor loadings from a model estimated with only a single latent factor. In all but two industries, the parameter estimates are consistent with the restrictions on the factor loadings imposed by our structural model. 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. These constraints on the factor loadings are satisfied in most industries.

In two industries, Plastics and Transport equipments, our estimates show that the labor variable responds less 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[16].

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 4, due to e.g. non-convex adjustment costs. The capital coefficient for each of the four equations in our system, (23), 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

[15] 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).

[16] Griliches and Hausman (1986) find an elasticity of labor to changes in output which is close to one when transitory changes in output were eliminated, while Biorn and Klette (1999) present lower estimates.

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 some of the mean reversion in the dynamic process for the observable variables[17].

7.2 Differences in efficiency across firms

7.2.1 Intrinsic differences are larger than the subsequent innovations

Table 4 presents various measures of the intrinsic differences relative to the differences in cumulated innovations within each of the six industries. Column 2 suggests that in all industries, the variance of the intrinsic efficiency differences accounts for the larger fraction of the non-transient firm heterogeneity. The fractions of the variance accounted for by the intrinsic differences vary from 53 percent in Radio/TV equipment (NACE 32) to 69 percent in Medical instruments (NACE 33). Our estimates of the variance accounted for by the cumulated, idiosyncratic innovations are based on the average firm age ($\overline{T}$) in each industry, which vary between 6.7 and 8.5 years (see Table 5).

Another way to express the same pattern is presented by the unconditional variance ratios in column 3, showing that the variance in intrinsic differences is between 1.2 and 2.3 times as large as the variance in the cumulated, idiosyncratic innovations.

However, these results do not 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 3 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 1995 are presented in column 5. The pattern from columns 3 and 4 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 are larger than the differences in the cumulated innovations in all six industries.

Table 4 shows that *both* intrinsic differences and differences in cumulated, idiosyncratic innovations are essential in all 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

[17] 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 non-convergence in the underlying, un-contaminated processes of interest. See also Quah (1993).

with 95 percent point-wise confidence intervals[18]. Both graphs in Figure 4 presents the average pattern across industries[19]. 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 second 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). Section 2 discussed how learning effects and depreciation can determine the distribution of innovations. The results in chart 1 suggests that young firms on average learn faster. According to our model, the high mean for the innovations shows that differences in performance tend be reinforced for young firms.

7.4 Systematic selection

The first chart of Figure 5 plots the estimated cohort-specific means of the intrinsic differences for all sectors combined, together with confidence intervals[20]. There is no systematic pattern suggesting e.g. that younger cohorts have particularly favorable intrinsic efficiency levels. On the other hand, in the second chart of Figure 5, we have plotted the mean of the intrinsic efficiency levels by the total life-time for each firm[21]. We overall find that firms with short life times (i.e. shorter than 5 years) have adverse intrinsic efficiency levels compared to firms with long life-times (i.e. more than 15 years). We interpret this as a mere self-selection effect: Intrinsic efficiency levels are positively correlated with survival. Hence, as $T_i - \tau_i$ increases we obtain an increasingly self-selected sample of firms.

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 pre-

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

[19]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.

[20]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.

[21]This graph is obviously contaminated by the problem that the life-time distribution of firms is right censored, but the pattern is still highly suggestive.

dicted 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 et al. (2001)[22]. However, our measurement of efficiency differs from the previous studies.

8 Conclusions

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 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.

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[23], 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.

References

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.

Biorn, 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 (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.

[22]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).

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

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.

Hall, B. (1987): The relationship between firm size and firm growth in the US manufacturing sector. *Journal of Industrial Economics*, 35, 583–606.

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. (2001): Comparing productivity at the firm level: What do we measure? Mimeo, University of Oslo.

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: Evidence and a theory. Mimeo. University of Oslo and Boston University.

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

Levinsohn, J. and M. Melitz (2001): Productivity in a differentiated products market equilibrium. Draft, Presented at the NBER Summer Institute, 2001.

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. (2001a): Identification, estimation and testing in panel data models with attrition: The role of the MAR assumption. Paper presented at the Annual Research Meeting for Norwegian Economists, 2001. ftp://ftp.ssb.no/pub/publikasjoner/DPxxx.pdf.

— (2001b): A state space approach for estimating VAR models for panel data with latent dynamic components. Discussion Paper 295, Statistics Norway. ftp://ftp.ssb.no/pub/publikasjoner/DP295.pdf.

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.

Wu, C. (1983): On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11, 95–103.

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[24]. The time-invariant part will be referred to as the intrinsic differences, while the cumulated part will be labelled as 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: How can we 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)[25]. 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 et al. (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 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.

[24] 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).

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

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[26].

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 Kortum (2001). 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

We assume that a_{it} is a random walk. However, from an econometric point of view, it might be desirable to generalize the latent process; for example by assuming that

$$s_{it} = a^*_{it} + d_t + e_{it}$$

where

$$a^*_{it} = \begin{cases} v_i & t = 1 \\ \phi a^*_{i,t-1} + \eta_{it} & t > 1 \end{cases}$$

Hence v_i is the intrinsic efficiency level whereas

$$a_{it} \equiv a^*_{it} - v_i$$

is the change in efficiency relative to the intrinsic level v_i. As before, we obtain the model

$$s_{it} = v_i + a_{it} + d_t + e_{it},$$

but the dynamics is different. Our assumption that $\phi = 1$ serves two purposes in our analysis. It is consistent with Gibrat's law, which has received some support in the empirical literature[27]. But perhaps

[26] 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.

[27] The empirical literature suggest that Gibrat's law is valid, when the inference doesn't condition on survival. See Hall (1987) and Klette and Kortum (2001) for a longer discussion and further references.

more importantly, it contributes to the identification of the model. The cost is that we fundamentally restrict the dynamics of the s_{it}-process. Some evidence, presented in Blundell and Bond (2000) indicate that s_{it} may be close to a unit root process, but that ϕ is still significantly lower than 1.

Our econometric procedure does not critically depend on the exact value of ϕ, and none of our results presented in section 7 would be seriously affected if ϕ is, in fact, slightly smaller than one. One reason for this is related to the role of the distribution of a_{it} in (2) as a smoothness-prior (see Kitagawa (1996) for a discussion of the role of the prior distribution of a stochastic trend in time series analysis). Moreover, we will argue that many of the empirical implications that can be deduced from $\phi < 1$ are consistent with $\phi = 1$ in the presence of self-selection. For example, negative correlation between a_{it} and v_i is implied by $\phi < 1$. But when $\phi = 1$, we would still expect that $Cov(a_{it}, v_i) < 0$ *conditional upon survival,* as argued in Section 7.4. Indeed, this negative relation is strongly confirmed by our data: The estimated correlation between a_{iT} and v_i (T corresponds to 1995) is on average -0.4, although the unconditional correlation, according to our model, is 0. Based on this and similar considerations, we claim that the models with $\phi = 1$ and ϕ near 1 may have more or less identical empirical implications, and therefore cannot meaningfully be distinguished based on these implications.

Appendix B: A simple model of capital accumulation

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

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

where $V(A_t^*, K_{t-1})$ is the value function and $E\left[\cdot|A_t^*\right]$ is the expectation conditional on A_t^*. A_t^* follows a 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

$$c'(K_t/K_{t-1}) = \frac{\partial}{\partial K_t}\beta E\left[V(A_{t+1}^*, K_t)|A_t^*\right]. \tag{36}$$

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 since $\Pi(A_t^*, K_{t-1}) - I(K_{t,}K_{t-1})$ is strictly concave in K_{t-1} and increasing in A_t^*, $E\left[V(A_{t+1}^*, K_t)|A_t^*\right]$ is strictly concave in K_t and increasing in A_t^*. It follows that the policy function is increasing in both arguments. Furthermore, it is concave in its first argument, which can be verified as follows. Define $\hat{k}_t = dK_t/K_t$, it follows from (36) 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:

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

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

$$\ln K_{it} = \alpha_t + \alpha_a \ln A_{it}^* + \alpha_k \ln K_{i,t-1}. \tag{37}$$

where α_a is positive while α_k is between zero and one. 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.

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: The ECM algorithm

The following factorization will be useful:

$$\Sigma_\eta = \Gamma_\eta \Gamma_\eta', \tag{38}$$

where Γ_v is a Cholesky-type $4 \times r$ matrix with zeros above the diagonal and positive diagonal elements. If $r = 4$, Γ_η is the unique Cholesky factor of Σ_η. In general, Γ_η is a unique rank-r decomposition of Σ_η.

We can now work with the following equivalent form of the state space representation (29):

$$\begin{aligned} \widetilde{\mathbf{y}}_{is} &= \Gamma_\eta \mathbf{G}_{is} \boldsymbol{\alpha}_{is} + \widetilde{\mathbf{d}}_{is} + \boldsymbol{\gamma}_k \widetilde{k}_{is} + \widetilde{\mathbf{R}}_{is} \\ \boldsymbol{\alpha}_{is} &= F \boldsymbol{\alpha}_{i,s-1} + \boldsymbol{\omega}_{is} \end{aligned} \qquad s = 1, ..., T_i - \tau_i + 1, \tag{39}$$

where the state vector $\boldsymbol{\alpha}_{is}$ has been standardized and no longer depends on unknown parameters:

$$\begin{aligned} \boldsymbol{\alpha}_{i0} &= \mathbf{0} \\ F &= \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{I}_r & \mathbf{I}_r \end{bmatrix} \\ \boldsymbol{\omega}_{is} &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}) \end{aligned}$$

where

$$\mathbf{Q} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

All other variables and parameters are defined in section 6.3.

The E-step: The E-step of the ECM algorithm now consists in evaluating

$$M(\boldsymbol{\beta}|\,\boldsymbol{\beta}^{(m)}) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{s=1}^{T_i-\tau_i+1}\left(\ \ln|\Xi_s| + \right. \tag{40}$$

$$\left. E\left\{\left(\widetilde{\mathbf{y}}_{is} - \Gamma_\eta \mathbf{G}_{is}\boldsymbol{\alpha}_{is} - \widetilde{\mathbf{d}}_{is} - \boldsymbol{\gamma}_k \widetilde{k}_{is}\right)' \Xi_s^{-1} \left(\widetilde{\mathbf{y}}_{is} - \Gamma_\eta \mathbf{G}_{is}\boldsymbol{\alpha}_{is} - \widetilde{\mathbf{d}}_{is} - \boldsymbol{\gamma}_k \widetilde{k}_{is}\right)\, |\mathbf{Y}^i; \boldsymbol{\beta}^{(m)}\right\}\right), \tag{41}$$

where

$$\Xi_s = \begin{cases} \Sigma_e & s = 1,..,T_i-\tau_i \\ \Sigma_v + s^{-1}\Sigma_e & s = T_i - \tau_i + 1 \end{cases}$$

(see equation (28)) and the expectation is with respect to $\boldsymbol{\alpha}_{is}$ given $\mathbf{Y}^i = \{\widetilde{\mathbf{y}}_{i1}, ..., \widetilde{\mathbf{y}}_{i,T_i-\tau_i+1}\}$.

Because (41) is the expectation of a function which is quadratic in $(\boldsymbol{\alpha}_{i1}, ..., \boldsymbol{\alpha}_{i,T_i-\tau_i+1})$, to evaluate this expectations we only need to calculate the conditional moments:

$$\begin{aligned} \mathbf{a}_{is|T_i-\tau_i+1} &= E\{\boldsymbol{\alpha}_{is}\,|\,\mathbf{Y}^i; \boldsymbol{\beta}^{(m)}\} \\ \mathbf{V}_{is\,|T_i-\tau_i+1} &= E\{(\boldsymbol{\alpha}_{is} - \mathbf{a}_{is|T_i-\tau_i+1})(\boldsymbol{\alpha}_{is} - \mathbf{a}_{is|T_i-\tau_i+1})'\,|\,\mathbf{Y}^i; \boldsymbol{\beta}^{(m)}\}. \end{aligned} \tag{42}$$

The state space form (39) can be used to derive (42) by means of the Kalman-filter and -smoother. By modifying the exposition in Fahrmeir and Tutz (1994), p. 264, the filtering recursions can be described by the following algorithm:

Kalman filtering:

$$\begin{aligned} &\text{For } i = 1, ..., N: \\ \mathbf{a}_{0|0} &= 0 \\ \mathbf{V}_{0|0} &= 0 \\ &\text{do for } s = 1, ..., T_i - \tau_i + 1: \\ \mathbf{a}_{is|s-1} &= F\mathbf{a}_{i,s-1|s-1} \\ \mathbf{V}_{is|s-1} &= F\mathbf{V}_{i,s-1|s-1}F' + \mathbf{Q} \\ \mathbf{G}^*_{is} &= \Gamma_\eta \mathbf{G}_{is} \\ \mathbf{Z}_{is} &= \widetilde{\mathbf{y}}_{is} - \widetilde{\mathbf{d}}_{is} - \gamma_k \widetilde{k}_{is} \\ \mathbf{K}_{is} &= \mathbf{V}_{is|s-1}\mathbf{G}^{*\prime}_{is}[\mathbf{G}^*_{is}\mathbf{V}_{is|s-1}\mathbf{G}^{*\prime}_{i} + \Xi_s]^{-1} \\ \mathbf{a}_{is|s} &= \mathbf{a}_{is|s-1} + \mathbf{K}_{is}(\mathbf{Z}_{is} - \mathbf{G}^*_{is}\mathbf{a}_{is|s-1}) \\ \mathbf{V}_{is|s} &= \mathbf{V}_{is|s-1} - \mathbf{K}_{is}\mathbf{G}^*_{is}\mathbf{V}_{is|s-1}, \end{aligned}$$

where all parameters are evaluated at $\boldsymbol{\beta} = \boldsymbol{\beta}^{(m)}$. The required conditional expectations $\mathbf{a}_{is|T_i-\tau_i+1}$ and variances $\mathbf{V}_{is\,|T_i-\tau_i+1}$ are obtained in subsequent backward smoothing recursions (see Fahrmeir and Tutz (1994), p. 265):

Kalman smoothing:

$$\begin{aligned} &\text{For } i = 1, ..., N: \\ &\text{do for } s = T_i - \tau_i + 1, ..., 2: \\ \mathbf{a}_{i,s-1|T_i-\tau_i+1} &= \mathbf{a}_{i,s-1|s-1} + \mathbf{B}_{is}(\mathbf{a}_{is|T_i-\tau_i+1} - \mathbf{a}_{is|s-1}) \\ \mathbf{V}_{i,s-1|T_i-\tau_i+1} &= \mathbf{V}_{i,s-1|s-1} + \mathbf{B}_{is}(\mathbf{V}_{is|T_i-\tau_i+1} - \mathbf{V}_{is|s-1)}\mathbf{B}'_{is}, \end{aligned}$$

where

$$\mathbf{B}_{is} = \mathbf{V}_{i,s-1|s-1}F'\mathbf{V}^{-1}_{is|s-1}.$$

The CM-step: In the m'th CM-step, after $\mathbf{a}_{is|T_i-\tau_i+1}$ and $\mathbf{V}_{is|T_i-\tau_i+1}$ have been evaluated in the preceding E-step, we update $\boldsymbol{\beta}$ to obtain $\boldsymbol{\beta}^{(m+1)}$. However, maximization of $M(\boldsymbol{\beta}|\boldsymbol{\beta}^{(m)})$ requires iterative methods. To simplify the calculations, we partition $\boldsymbol{\beta}$. Let $\boldsymbol{\beta}_1 = (\Gamma_\eta, \boldsymbol{\gamma}_k, \mathbf{d})$, $\boldsymbol{\beta}_2 = (\Sigma_v, \Sigma_e)$, and $M(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2|\boldsymbol{\beta}) \equiv M(\boldsymbol{\beta}|\boldsymbol{\beta})$. The CM-step consists of the conditional maximizations:

$$\boldsymbol{\beta}_1^{(m+1)} = \arg\max_{\boldsymbol{\beta}_1} M(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2^{(m)}|\boldsymbol{\beta}^{(m)}) \tag{43}$$

$$\boldsymbol{\beta}_2^{(m+1)} = \arg\max_{\boldsymbol{\beta}_2} M(\boldsymbol{\beta}_1^{(m+1)}, \boldsymbol{\beta}_2|\boldsymbol{\beta}^{(m)}). \tag{44}$$

Conditional maximization w.r.t. $\boldsymbol{\beta}_1$ in (43) is equivalent to minimization of the quadratic function

$$q(\boldsymbol{\beta}_1) = \sum_{i=1}^{N} \sum_{s=1}^{T_i-\tau_i+1} \left(\widetilde{\mathbf{y}}_{is} - \Gamma_\eta \mathbf{G}_{is}\mathbf{a}_{is|T_i-\tau_i+1} - \widetilde{\mathbf{d}}_{is} - \boldsymbol{\gamma}_k \widetilde{k}_{is}\right)' \Xi_s^{-1} \left(\widetilde{\mathbf{y}}_{is} - \Gamma_\eta \mathbf{G}_{is}\mathbf{a}_{is|T_i-\tau_i+1} - \widetilde{\mathbf{d}}_{is} - \boldsymbol{\gamma}_k \widetilde{k}_{is}\right)'$$
$$+ \sum_{i=1}^{N} \sum_{s=1}^{T_i-\tau_i+1} \text{tr } \Xi_s^{-1} \Gamma_\eta \mathbf{G}_{is} \mathbf{V}'_{is|T_i-\tau_i+1} \mathbf{G}'_{is} \Gamma_\eta{}' .$$

By setting the derivative of $q(\boldsymbol{\beta}_1)$ equal to zero, we get linear 1. order conditions. The updating of $\boldsymbol{\beta}_1$ in (43) is therefore trivial.

Maximization with respect to $\boldsymbol{\beta}_2$ in (44) is equivalent to minimization of

$$r(\boldsymbol{\beta}_2) = \sum_{i=1}^{N} \sum_{s=1}^{T_i-\tau_i+1} \left(\ln|\Xi_s| + \text{tr } \Xi_s^{-1} \left(\mathbf{S}_{is}\mathbf{S}_{is}{}' + \Gamma_\eta \mathbf{G}_{is}\mathbf{V}_{is|T_i-\tau_i+1}\mathbf{G}_{is}{}'\Gamma_\eta{}'\right)\right) \tag{45}$$

where

$$\mathbf{S}_{is} = \widetilde{\mathbf{y}}_{is} - \Gamma_\eta \mathbf{G}_{is}\mathbf{a}_{is|T_i-\tau_i+1} - \widetilde{\mathbf{d}}_{is} - \boldsymbol{\gamma}_k \widetilde{k}_{is}$$

Note that $r(\boldsymbol{\beta}_2)$ depends on $\boldsymbol{\beta}_2 = (\Sigma_v, \Sigma_e)$ only through Ξ_s. Rather than expressing Ξ_s as functions of (Σ_v, Σ_e), it is useful to write

$$\begin{aligned} \Sigma_e &= \Gamma_e \Gamma_e{}' \\ \Sigma_v &= \Gamma_v \Gamma_v{}' \end{aligned} \tag{46}$$

where Γ_e is the 4×4 lower triangular Cholesky factor of Σ_e, and Γ_v is a Cholesky-type $4 \times r$ matrix with zeros above the diagonal and positive diagonal elements.

Through the reparametrizations (46), we assure that the estimates of the covariance matrices are positive semi-definite and that the estimate of Σ_v has rank r. Unfortunately there is no closed-form solution to the problem of minimizing $r(\boldsymbol{\beta}_2)$ with respect to the (unique) Cholesky (-type) factors (Γ_e, Γ_v). However, analytic expressions for the derivatives of the objective function (45) with respect to the components of Γ_e and Γ_v are available (see Lutkepohl (1996)), and in our experience this optimization problem is easily solved numerically, by using an efficient quasi-Newton algorithm.

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 (11)-(12), and define $\overline{\eta}_s \equiv E(\eta_{i,\tau_i+s}^*)$ – the expected innovation of a firm of age s. Then $\overline{\eta}_s$ should decrease as s increases.

Consider the *a posteriori* distribution of the innovations of firms with age s, i.e. η_{i,τ_i+s}. That is, the distribution of η_{i,τ_i+s} conditional on all the observations on firm i (evaluated at the parameter estimate

$\widehat{\boldsymbol{\beta}}$). Let $\widehat{\eta}_{it}$ be the Kalman-smoothed estimate of η^*_{it}: $\widehat{\eta}_{it} = E(\eta^*_{it}|\text{all the data on firm } i)$. Then we can estimate $\overline{\eta}_s$ using the sample analogy method:

$$\overline{\eta}_s \approx \frac{1}{n_s} \sum_{i \in N_s} \widehat{\eta}_{i,\tau_i+s},$$

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

For fixed s, $Var(\eta^*_{i,\tau_i+s}) = Var(\widehat{\eta}_{i,\tau_i+s}) + E(Var(\widehat{\eta}_{i,\tau_i+s}))$. 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 $\widehat{\eta}_{i,\tau_i+s}$ and $Var(\widehat{\eta}_{i,\tau_i+s})$, which are outputs from the Kalman-smoother (cf. section 6.2).

The relation between intrinsic differences and selection in section 7.3 is studied from the smoothed innovations $\widehat{\mathbf{v}}_i = E(\mathbf{v}_i|\text{all the data on firm } i)$. 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 (13).

Table 1: **Descriptive statistics**

Sector	NACE	#Firms	# Firms in 95	Mean sales (Std.)	Median sales	Lab.prod. (Std)
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)

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)	**Psuedo 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 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. (γ_k)	Psuedo 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: **Importance of cumulative shocks vs. inital conditions and their correlation**

Sector (NACE)	$\frac{\sigma_v^2}{\sigma_v^2+\sigma_\eta^2 T}$	$\frac{\sigma_v^2}{\sigma_\eta^2 T}$	$\frac{tr\ Var(v_i \mid i \in M_T)}{tr\ Var(a_{it} \mid i \in M_T)}$	$Corr(a_{it}, v_i \mid i \in M_T)$
Plastics (25)	.66	2.0	2.3	$-.27$
Machinery (29)	.54	1.2	1.7	$-.39$
Electrical inst. (31)	.54	1.2	1.2	$-.55$
Radio/ TV eq. (32)	.53	1.2	2.0	$-.44$
Medical inst. (33)	.69	2.3	2.6	$-.50$
Transp. eq. (35)	.68	2.1	2.6	$-.34$

Table 5: **Average firm age, and the variances of cumulative innovations and intrinsic differences**

Sector (NACE)	**Avg. age**	σ_η^2	σ_v^2	$T^* = \frac{\sigma_v^2}{\sigma_\eta^2}$
Plastics (25)	7.1	0.16	2.27	14.2
Machinery (29)	6.9	0.20	1.66	8.3
Electrical inst. (31)	7.2	0.20	1.80	9.0
Radio/ TV eq. (32)	8.5	0.32	3.20	10.0
Medical inst. (33)	6.7	0.23	3.46	15.0
Transp. eq. (35)	8.5	0.24	4.25	17.7

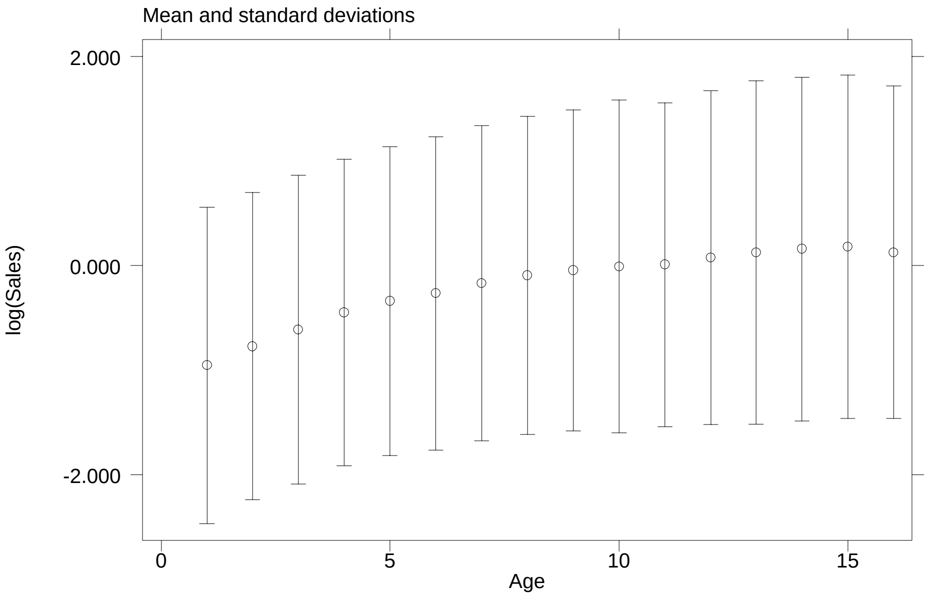


Figure 1: **Differences in log sales as a function of firm age.** Circles indicate the means and whiskers show the standard errors.

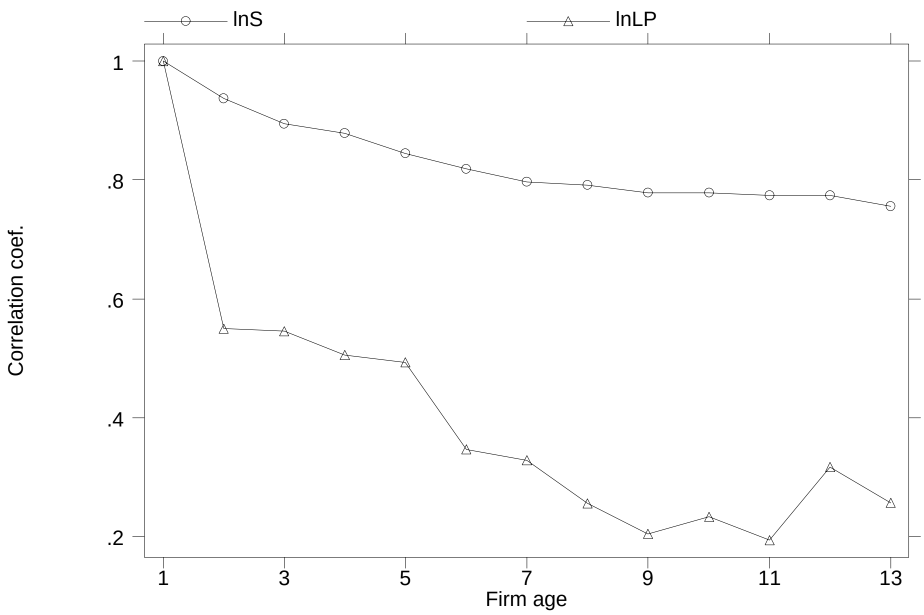


Figure 2: **The correlation between relative performance in a firm's first year and in its subsequent years.** The circles correspond to the correlation coefficents for (log) sales while the triangles refer to (log) labor productivity.

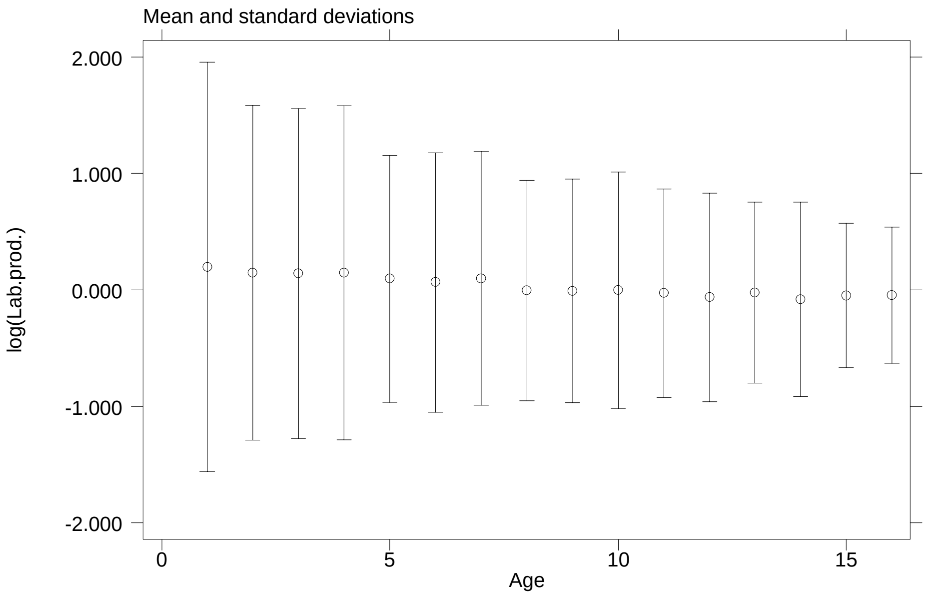


Figure 3: **Differences in log labor productivity as a function of firm age.** Circles indicate the means and whiskers show the standard errors.

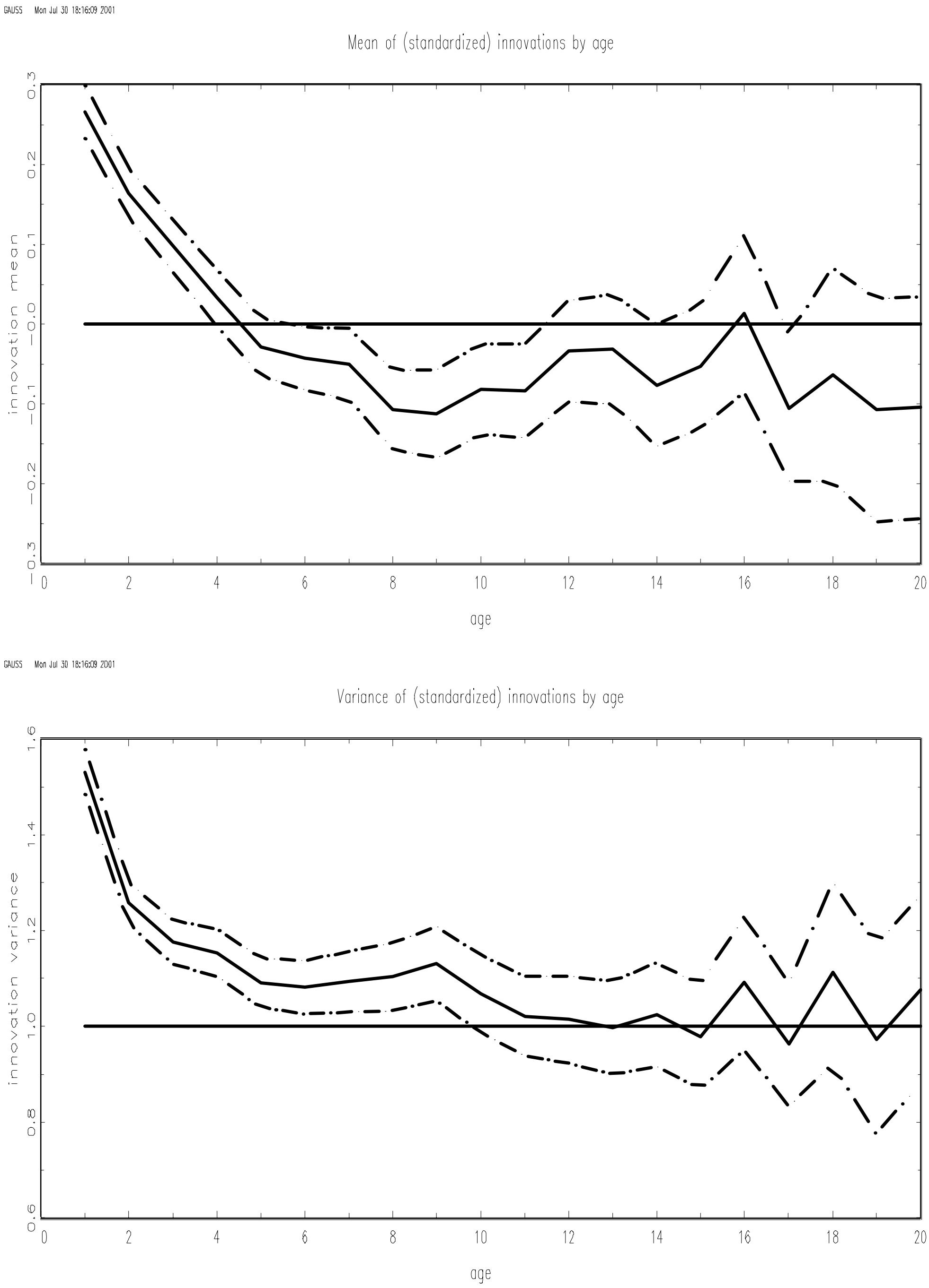
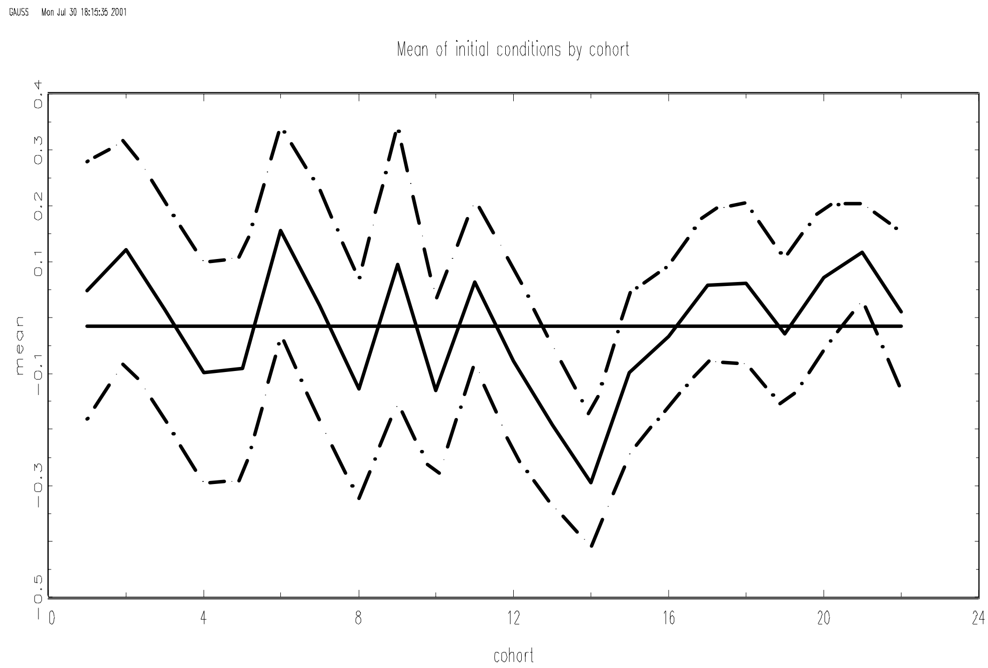


Figure 4: **The mean (upper chart) and variance (lower chart) of the innovations decrease with the age of the firms.** Standard deviations indicated by dotted lines.



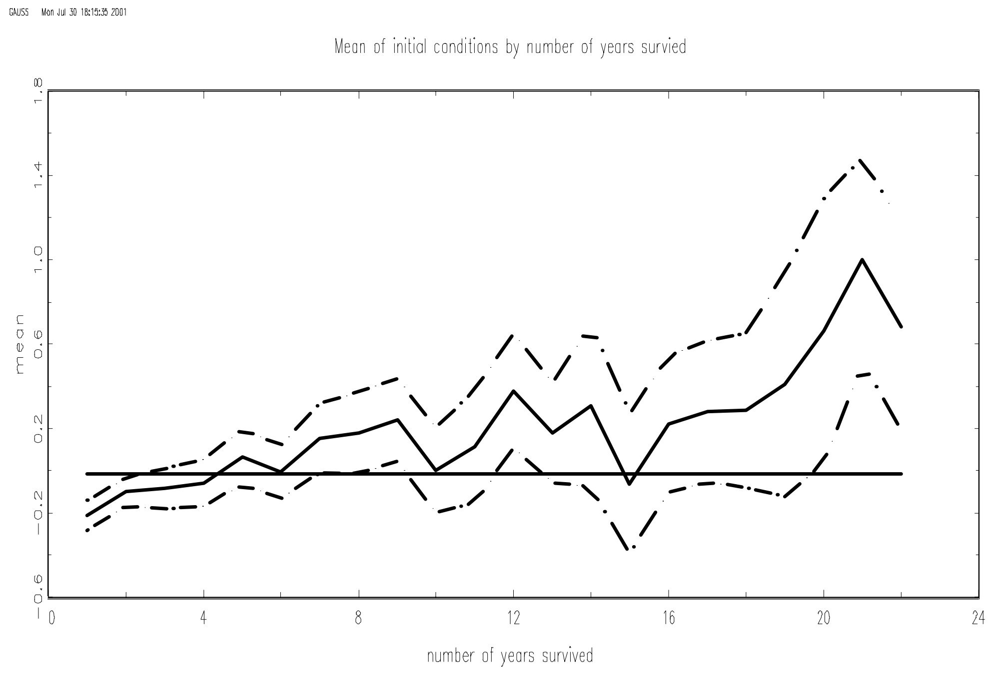


Figure 5: **No systematic differences across cohorts in initial conditions (upper chart). Firms with higher initial productivity live longer (lower chart).** Standard deviations indicated by dotted lines.

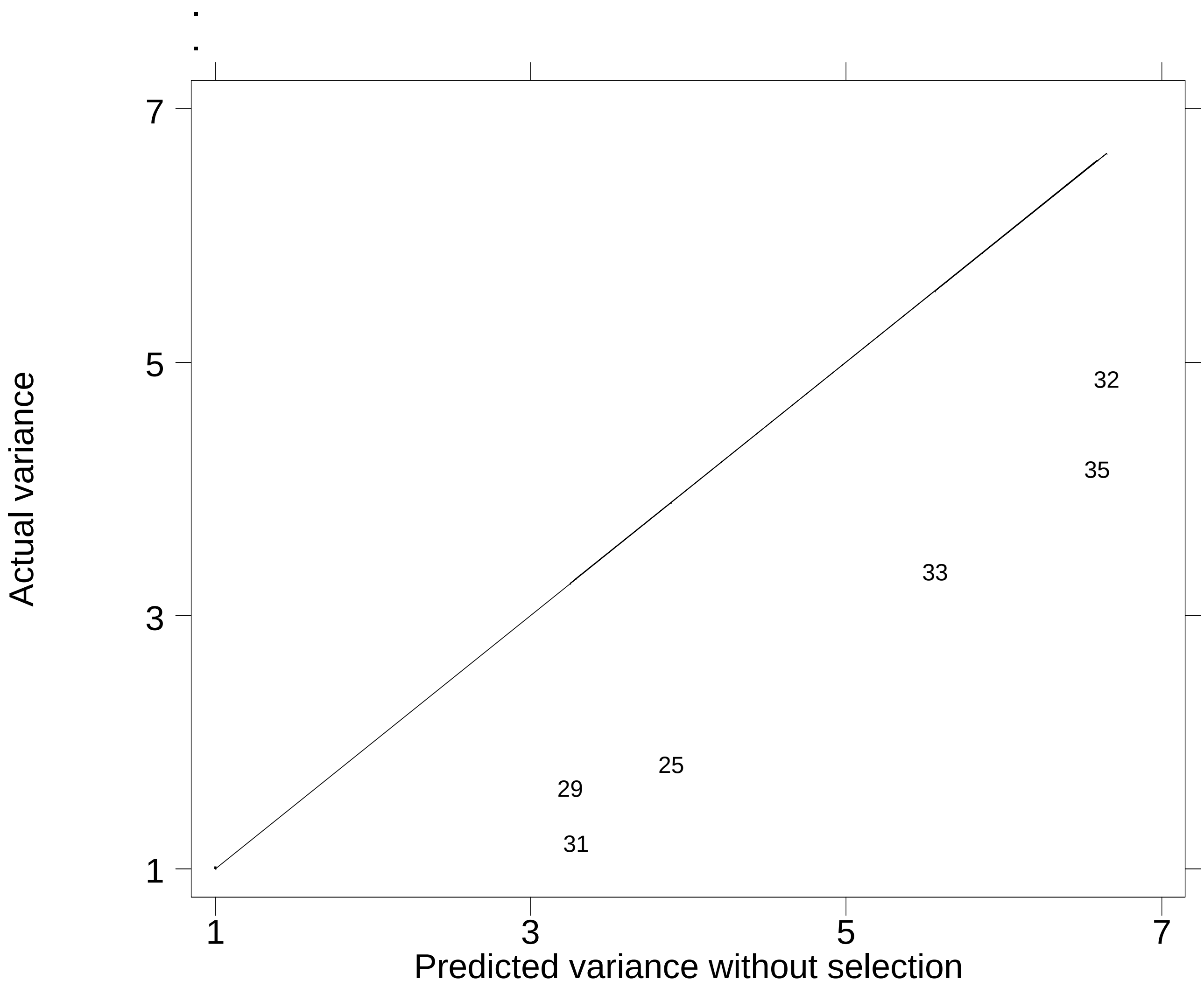


Figure 6: **Observed efficiency differences much smaller than predicted efficiency differences in the absence of selection.** Variances of the observed efficiency differences on the vertical axis and predicted efficiency differences in the absence of selection on the horizontal axis. 45-degree line also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