Tor Jakob Klette and Arvid Raknerud

Heterogeneity, productivity and selection: an empirical study of Norwegian manufacturing firms

Abstract:
How do firms differ, and why do they differ even within narrowly defined industries? Using evidence from a new panel data set for four high-tech, manufacturing industries covering a 10-year period, we show how differences in sales, materials, labor costs and capital across firms can be summarized by firm-specific, dynamic factors, which we interpret in view of a structural model. The model contains the complete system of supply and factor demand equations. Our results show that a firm's efficiency is strongly linked to profitability and firm size, but only weakly related to labor productivity. Our second task is to understand the origin and evolution of the differences in efficiency. Among the firms established within the 10-year period that we consider permanent differences in efficiency dominate over differences generated by firm-specific, cumulated innovations.

1 Introduction

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

Efficiency differences are decomposed into stochastic, firm-specific (idiosyncratic) *cumulated innovations* as emphasized e.g. by Ericson and Pakes (1995), and *permanent efficiency differences* as emphasized by Jovanovic (1982) and others[1]. In the four high-tech industries that we examine, the efficiency differences are largely permanent, while cumulated innovations play a lesser role.

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

We use the term efficiency rather than productivity, since economic theory does not predict an unambigous relationship between differences in labor productivity and differences in efficiency: Consider firms with different levels of efficiency competing in a frictionless industry. A firm with high efficiency will choose a high level of labor input so that its marginal product is equal to the real wage, which, by assumption, is the same across all firms[2]. With a Cobb-Douglas production function, the marginal product is proportional to production per factor input, and, hence, all firms should have the same level

---

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

of production per factor input apart from transient noise or fluctuations. This argument raises the question of how to make inferences about differences in efficiency from firm level data, which is a central theme in our analysis.

Our empirical framework decomposes the observations of firm-level supply and factor demands in terms of four types of latent components: (i) firm-specific permanent components, (ii) firm-specific stochastic trends, (iii) transient noise, and (iv) industry-wide fluctuations. The multivariate framework imposes few restrictions on the data generating process *a priori* and allows us to consider the validity of restrictions imposed by formal economic models. Our empirical model explicitly accounts for sample self-selection caused by firm exit.

2 A first look at differences in firm performance

How should one measure differences in firm performance, and do these differences increase with firm age? Using size as a measure of firm performance, we approach the second question by examining Figure 1[3]. The figure presents the mean and interquartile range of log sales as a function of firm age among firms established during 1993-2002. All observations are measured relative to industry-year averages at the two digit NACE level. Not surprisingly, the figure shows that young firms are substantially smaller than older firms: the average size of eight year old firms is 25 percent lower than the industry average, and firm growth tends to decelerate with age. More interestingly, the figure shows that the interquartile range is almost independent of firm age, indicating that the magnitude of firm heterogeneity is unrelated to size.

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 for the first and second year is 0.92, and it declines slowly during the subsequent years. This shows that the *relative* differences in firm size are highly persistent as firms become older.

These patterns indicate that differences across young firms are as large as those among older firms and that the differences are highly persistent. However, this conclusion is preliminary as it leaves open a number of questions. Young firms have a high rate of

---

[3]Figures 1-2 are based on a comprehensive, unbalanced sample of firm level observations from four high-tech manufacturing industries, as discussed in Section 6.

exit; on average, 40 percent of a new cohort of firms have exited within eight years in our sample. Since exiting firms are systematically selected among the least successful firms, we expect an upward trend in average log sales, as confirmed by Figure 1. Systematic selection that eliminates the least successful firms should also, *cet.par.*, tend to narrow down the differences in firm size. However, there is no evidence of this in the figure. There seems to be an offsetting force that tends to make firms grow more unequal with age. Such an offsetting force could be idiosyncratic, cumulated shocks that would also explain the declining correlation between a firm's performance in its first year and in its subsequent years, as depicted in Figure 2.

Labor productivity is another widely used measure of firm performance. Figure 3 presents the mean and interquartile range of log labor productivity as a function of firm age, where labor productivity is measured as value added per man hour[4]. We see that the patterns are rather different from those in Figure 1. There is no upward trend in labor productivity and the interquartile range declines with age. The difference between sales and labor productivity is equally clear from the lower graph in Figure 2, which displays the correlation coefficient between labor productivity in the firms' first year and in their subsequent years[5]. The correlation coefficient between productivity in the first and second year is only 0.40, which indicate that 60 percent of the observed variance in log labor productivity is due to temporary fluctuations or noise in the data. A comparison of the two graphs in Figure 2 raises the question of why differences in size are considerably more persistent than differences in labor productivity.

The results in Figure 1 and 2 are in good agreement with Baumol and Wolff (1984), who showed that many commonly used output-input ratios fail to pick up efficiency differences in a competitive equilibrium where all firms have zero profit. However, we seem to be at odds with a host of empirical studies on micro data; e.g. Baily, Hulten and Campell (1992), who studied U.S. manufacturing data; and more recently Foster, Haltiwanger and Kirzan (2002), who examined U.S. retail trade data. These, and many other papers, report a high degree of persistence in measured labor productivity.

---

[4]Value added is a real income measure, defined as $(p_t Q_t - w_t^m M_t)$, where p_t is the output price, w_t^m is the price of material inputs, Q_t is output, and M_t is material inputs. All prices are real prices obtained by applying a common deflator.

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

We suspect that the discrepancies between different empirical studies is due mainly to the choice of measurement method: While we use man hours, most published studies use number of employees as (a proxy for) labor input. In fact, data on sales per employee shows that Norwegian data are *not* atypical: the correlation coefficient in log sales per employee between the firm's first and second year (subtracting annual industry-wide effects) is 0.77, decreasing to 0.51 after eight years[6]. This suggests that a firm's adjustment of number of employees is much slower than its adjustment of man hours.

3 A structural model of firm behavior

Section 3.1 presents a simple model of supply and factor demand based on standard assumptions of price taking behavior and Hicks-neutral technical change. This model is extended in Section 3.2, to allow heterogeneous prices and labor augmenting innovations. The structural models provide the theoretical foundation for the econometric framework that we later use to make inferences about unobserved differences in efficiency from observed sales and factor costs.

3.1 A model of firm supply and factor demand

Assume that firm i has the production function:

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

where Q_{it} and A_{it} denote firm i's output and efficiency in year t, $K_{i,t-1}$ is the predetermined capital stock at the beginning of year t (i.e. end of $t-1$), and $F(M_{it}, L_{it})$ is a function aggregating the two fully flexible production factors materials, M_{it}, and labor inputs, L_{it}. The aggregation function is homogenous of degree ε, with elasticity of scale $\gamma + \varepsilon < 1$. Given that the firms face the same price of output, p_t, materials, w_t^m, and labor, w_t^l, it follows that the short-run cost-function has the form:

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

where $c(W_t)$ is a function that is homogeneous of degree one in $W_t \equiv \left(w_t^l, w_t^m\right)$. Setting price equal to marginal costs, we obtain the following set of equations for sales, $S_{it} \equiv p_t Q_{it}$,

---

[6] If value added is used as output measure, the corresponding figures become 0.58 and 0.50, respectively.

and short-run factor demand:

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

where $\mathbf{g}(P_t)$ is a vector function that only depends on the price vector $P_t \equiv (p_t, W_t)$. The resulting short-run profit function is homogeneous in capital:

$$
\begin{aligned}
\Pi(A_{it}, K_{i,t-1}, P_t) &= \pi(P_t) \left(A_{it} K_{i,t-1}^{\gamma}\right)^{1/(1-\varepsilon)} \\
&\equiv \Pi_{it} K_{i,t-1}^{\gamma/(1-\varepsilon)}. \qquad (4)
\end{aligned}
$$

According to (3), differences in firm output, material use and labor input are informative about *unobserved* differences in firm efficiency, conditional on the firms' capital stocks. However, the equations in (3) cannot be directly exploited to make inferences about the differences in efficiency, as these tend to be (positively) correlated with differences in capital.

Capital stock dynamics: We shall next consider the capital stock dynamics derived from assumptions about firms' investment behavior. Assume that the multiplicative factor Π_{it} in (4) is a Markovian stochastic process and that the adjustment costs of capital are weakly convex due to partial irreversibilities[7]. Then we can use results from Bloom (2000) and Bloom et al. (2001), who demonstrate that the actual capital stock at the beginning of year t, $K_{i,t-1}$, and the hypothetical frictionless capital stock, $K_{i,t-1}^*$, will have the same long run growth rate. That is:

$$
\ln K_{i,t-1} = \ln K_{i,t-1}^* + \text{error}, \qquad (5)
$$

where the error term is stationary. To be more specific, $K_{i,t-1}^*$ is the capital stock the firm would choose if the marginal revenue of capital, i.e. the derivative of $\Pi(A_{it}, K_{i,t-1}, P_t)$ with respect to $K_{i,t-1}$, is equal to the Joergensonian user cost.

It is easy to see from (4) that $\ln K_{i,t-1}^*$ must be linear in $\ln A_{it}$. A first order approximation to a general equilibrium correction model is therefore:

---

[7]Adjustment costs may also apply to labor input. However, there can be no doubt that labor is a much more flexible production factor than fixed capital, which, due to large transaction costs and lack of well-functioning second hand markets for many types of capital, often have low alternative value outside its current use or location. In contrast, the costs of adjusting man hours are comparably small in the Norwegian labor market.

$$\Delta \ln K_{it} = (\phi - 1)(\ln K_{i,t-1} - \kappa_A \ln A_{it}) + \kappa_t, \tag{6}$$

where $|\phi| < 1$, κ_A is a fixed parameter and κ_t is an industry-wide time varying intercept[8].

Even in the presence of kinks in the adjustment cost function due to partial irreversibilities, investments tend to be relatively smooth at the firm level when only one type of aggregate capital is considered. This is well documented both on Norwegian and international data sets[9] and motivates a smooth error correction model like (6) to describe the short-run dynamics of capital formation.

Supply and factor demand: Combining (3) and (6), we obtain a simultaneous system of equations:

$$\mathbf{y}_{it} = \boldsymbol{\theta}_A \ln A_{i1} + \boldsymbol{\theta}_A \ln(A_{it}/A_{i1}) + \boldsymbol{\theta}_K \ln (K_{i,t-1}) + \boldsymbol{\theta}_t, \tag{7}$$

where

$$
\begin{aligned}
\mathbf{y}_{it} &\equiv \left[\begin{array}{cccc} \ln S_{it}, & \ln M_{it}, & \ln L_{it}, & \ln K_{it} \end{array}\right]' \\
\boldsymbol{\theta}_A &= \left[\begin{array}{cccc} \frac{1}{1-\varepsilon}, & \frac{1}{1-\varepsilon}, & \frac{1}{1-\varepsilon}, & (1-\phi)\kappa_A \end{array}\right]' \\
\boldsymbol{\theta}_K &= \left[\begin{array}{cccc} \frac{\gamma}{1-\varepsilon}, & \frac{\gamma}{1-\varepsilon}, & \frac{\gamma}{1-\varepsilon}, & \phi \end{array}\right]' \\
\boldsymbol{\theta}_t &= \left[\begin{array}{cc} \mathbf{g}(P_t)', & \kappa_t \end{array}\right]'.
\end{aligned} \tag{8}
$$

The structural model (7)-(8) suggests that *differences* between firms in the endogenous variables, $\mathbf{y}_{it}$, are due to differences in *efficiency*, $\ln(A_{it})$, and *capital accumulation*, $\ln(K_{i,t-1})$. Equation (7) decomposes differences in efficiency into two components: permanent differences already introduced when the firms are established, $\ln A_{i1}$, and differences in subsequent innovations, i.e. the cumulated changes in efficiency, $\ln(A_{it}/A_{i1})$.

3.2   An extended model: Idiosyncracies in prices and labor productivity.

The model (7)-(8) puts heavy constraints on the data as it assumes that efficiency changes affect all the components of $\mathbf{y}_{it}$ through a single latent variable, A_{it}, and that

---

[8]Much of the recent theory on investment behavior focuses either on partial irreversibilities; i.e. the resale price of capital is lower than the purchasing price (e.g. Caballero et al., 1995; Abel and Eberly, 1996) or on fixed adjustment costs (e.g. Caballero and Engel, 1999). The model of Bloom (2000) builds on the first strand of this literature, allowing weakly convex adjustment costs that are kinked at zero due to partial irreversibilities.

[9]See Bond et al. (2001) and Nilsen and Schiantarelli (2003).

the three first components of the loading vector, $\boldsymbol{\theta}_A$, are equal. Moreover, according to (4), short-run profitability is increasing in efficiency A_{it} and capital $K_{i,t-1}$. On the other hand, (3) shows that differences in labor productivity, i.e. value added per labor input $(p_t Q_{it} - w_t^m M_{it})/L_{it}$, are independent of differences in firm efficiency, A_{it}. The last implication is, of course, not invariant with respect to the choice of production function.

Although common in empirical work, the production function (1) and the assumption of price taking behavior are quite restrictive. We will therefore consider an alternative specification, which will lead to an extension of the system of supply and factor demand (7). This specification incorporates heterogeneous prices as well as biased technical change.

Assume monopolistic competition between a large number of producers of a differentiated good, where each producer faces a demand function of the form:

$$Q_{it} = \Phi_{it} p_{it}^{-e},$$

where Φ_{it} is a demand shift parameter and e is the demand elasticity, $e > 1$. Each firm has a production function of the type (1), except that the aggregation function $F(\cdot, \cdot)$ is firm-time specific, denoted $F_{it}(\cdot, \cdot)$. To be able to identify the latent variables of the model from data on sales and factor demand, we shall confine the analysis to a special case:

$$F_{it}(M_{it}, L_{it}) = [(b_{it} L_{it})^\rho + M_{it}^\rho]^{\frac{\epsilon}{\rho}}, \text{ with } \rho < 1. \tag{9}$$

That is, a CES function with a labor augmenting parameter b_{it} and substitution parameter ρ. The corresponding cost function is:

$$C_{it}(Q_{it}, K_{i,t-1}) = c_{it} \left(\frac{Q_{it}}{A_{it} K_{i,t-1}^\gamma} \right)^{1/\varepsilon} \tag{10}$$

where

$$c_{it} = \left[(w_t^l/b_{it})^{\frac{\rho}{\rho-1}} + (w_t^m)^{\frac{\rho}{\rho-1}} \right]^{\frac{\rho-1}{\rho}}.$$

The profit maximizing system of sales and (short-run) factor demand equations (ignoring additive constants) is:

$$\begin{bmatrix} \ln S_{it} \\ \ln M_{it} \\ \ln L_{it} \end{bmatrix} = \begin{bmatrix} \frac{e-1}{\varepsilon+e-e\varepsilon} & -\frac{\varepsilon(e-1)}{\varepsilon+e-e\varepsilon} & 0 \\ \frac{e-1}{\varepsilon+e-e\varepsilon} & -\frac{\varepsilon(e-1)}{\varepsilon+e-e\varepsilon} + \frac{\rho}{1-\rho} & 0 \\ \frac{e-1}{\varepsilon+e-e\varepsilon} & -\frac{\varepsilon(e-1)}{\varepsilon+e-e\varepsilon} + \frac{\rho}{1-\rho} & \frac{\rho}{1-\rho} \end{bmatrix} \begin{bmatrix} \ln A_{it}^* \\ \ln c_{it} \\ \ln b_{it} \end{bmatrix} + \mathbf{1} \frac{\gamma(e-1)}{\varepsilon+e-e\varepsilon} \ln K_{i,t-1}, \tag{11}$$

where

$$A_{it}^* = \Phi_{it}^{1/(e-1)} A_{it}.$$

The system (11) has three linearly independent factors; $\ln A_{it}^*$, $\ln c_{it}$ and $\ln b_{it}$. The latter two factors will be negatively correlated, since c_{it} – which can be interpreted as a variable factor price index – is a monotonically decreasing function of b_{it} for given factor prices (w_t^l, w_t^m). We see that labor productivity will depend on $\ln c_{it}$ and $\ln b_{it}$, but not on $\ln A_{it}^*$.

The main structure of our demand and supply equations is not affected by the assumption of monopolistic competition: Efficiency and demand changes, due to e.g. quality differences, enter the system of equations (11) in an entirely symmetric way through the variable A_{it}^*. Consequently, we are not able to distinguish between these two types of shocks in the empirical analysis. We may think of A_{it}^* as as efficiency in a wide-sense, as it incorporates demand idiosyncracies. Since

$$\frac{(e-1)}{\varepsilon + e - e\varepsilon} \rightarrow \frac{1}{1-\varepsilon} \quad \text{and} \quad A_{it}^* \rightarrow A_{it} \text{ when } e \rightarrow \infty,$$

a model with price taking firms is obtained as a limiting case of (11).

The effects of a labor augmenting innovation, i.e. a positive chock in b_{it}, depends on the substitution parameter ρ. There are three main cases: If $\rho < 0$, a positive shock in b_{it} will lead to an increase in the ratio of materials to labor input. That is to say, the innovation is "labor saving" (see Binswanger, 1974). On the other hand, if $\rho > 0$, a labor augmenting innovation leads to a more labor intensive production. Finally, when $\rho = 0$ (or formally: $\rho \rightarrow 0$) the loading coefficient of $\ln b_{it}$ becomes zero, while c_{it} becomes a Cobb-Douglas function in the two arguments q_L/b_{it} and q_{M_t}. Thus b_{it} reduces to a Hicks-neutral efficiency shock indistinguishable from A_{it}.[10]

The short run profit function corresponding to (11) is homogeneous in capital. Furthermore, it is easy to cheque that frictionless capital, $K_{i,t-1}^*$, has the form:

$$\ln K_{i,t-1}^* = \kappa_A \ln A_{it}^* + \kappa_c \ln c_{it} + \kappa_t,$$

for fixed parameters κ_A and κ_c and a time varying intercept κ_t. Therefore, using (5) in combination with a first order error correction model, as in (6), the complete system of

---

[10]In the limiting case it is useful to include a share parameter, ν, in the CES function: $F_{it}(M_{it}, L_{it}) = [\nu(b_{it}L_{it})^\rho + (1-\nu)M_{it}^\rho]^{\frac{\varepsilon}{\rho}} \rightarrow (b_{it}L_{it})^{\nu\varepsilon} M_{it}^{(1-\nu)\varepsilon}$

equations for the extended model can be written:

$$\mathbf{y}_{it} = \boldsymbol{\theta}_A \, \boldsymbol{\alpha}_{i1} + \boldsymbol{\theta}_A \left(\boldsymbol{\alpha}_{it} - \boldsymbol{\alpha}_{i1} \right) + \boldsymbol{\theta}_K \ln \left(K_{i,t-1} \right) + \boldsymbol{\theta}_t \tag{12}$$

where

$$
\begin{aligned}
\boldsymbol{\alpha}_{it} &= \begin{bmatrix} \ln A_{it}^* & \ln c_{it} & \ln b_{it} \end{bmatrix}' \\
\boldsymbol{\theta}_A &= \begin{bmatrix}
\frac{e-1}{\varepsilon+e-e\varepsilon} & -\frac{\varepsilon(e-1)}{\varepsilon+e-e\varepsilon} & 0 \\
\frac{e-1}{\varepsilon+e-e\varepsilon} & -\frac{\varepsilon(e-1)}{\varepsilon+e-e\varepsilon} + \frac{\rho}{1-\rho} & 0 \\
\frac{e-1}{\varepsilon+e-e\varepsilon} & -\frac{\varepsilon(e-1)}{\varepsilon+e-e\varepsilon} + \frac{\rho}{1-\rho} & \frac{\rho}{1-\rho} \\
(1-\phi)\kappa_A & (1-\phi)\kappa_c & 0
\end{bmatrix} \\
\boldsymbol{\theta}_K &= \begin{bmatrix} \frac{\gamma(e-1)}{\varepsilon+e-e\varepsilon}, & \frac{\gamma(e-1)}{\varepsilon+e-e\varepsilon}, & \frac{\gamma(e-1)}{\varepsilon+e-e\varepsilon}, & \phi \end{bmatrix}'
\end{aligned}
\tag{13}
$$

and $\boldsymbol{\theta}_t$ is a time-varying intercept vector.

Without further restrictions, it is not possible to disentangle the effects of Hicks-neutral efficiency or demand shocks on the one hand and the effects of labor augmenting innovations on the other. The reason is that if the vector of latent factors $\boldsymbol{\alpha}_{it}$ in (12) is premultiplied with any 3×3 matrix $\mathbf{R}$, we obtain an observationally equivalent model by postmultiplying the loading coefficient matrix $\boldsymbol{\theta}_A$ with $\mathbf{R}^{-1}$. As we discuss in Section 4, one restriction is needed in addition to the particular structure of $\boldsymbol{\theta}_A$ in (13) to obtain identification: namely that the innovations $\Delta \ln A_{it}^*$ and $(\Delta \ln c_{it}, \Delta \ln b_{it})$ are mutually independent vectors.

A main feature of the extended model (12) is that it may be able to account for persistent of labor productivity differences, as exhibited in Figure 2. There are, of course, other explanations for such differences. One possibility is the presence of overhead labor. That is, each firm has a minimum amount of labor which is necessary in order to operate, regardless of the level of output; see Aghion and Howitt (1994). Overhead labor creates productivity differences, because labor inputs below the threshold have zero marginal product. One can show that labor productivity will be positively correlated with efficiency and negatively correlated with the threshold. While the assumption of overhead labor may motivate the use of labor productivity as a measure (or proxy) for efficiency; as in Haltiwanger et al. (2002), the explanatory power of this theory is weak when overhead labor makes up a small share of total labor input. In fact, there are reasons to believe that this is the case in manufacturing: When sales increase in the presence of overhead labor, we would expect an increase in labor productivity due to a positive scale effect.

11

However, no such pattern is visible in Figure 3, contrary to what we would expect from the positive relation between firm size and firm age depicted in Figure 1.

In the next section we will formulate a general econometric model that encompasses the simple model (7)-(8) with price taking behavior, as well as the extended model (11).

4   The econometric model

In this section we formulate our econometric model. This model imposes fewer restrictions on the data generating process than our structural models. We assume that:

$$\mathbf{y}_{it} = \mathbf{v}_i + \mathbf{a}_{it} + \boldsymbol{\theta}_K \ln K_{i,t-1} + \boldsymbol{\theta}_t + \mathbf{e}_{it}, \ \ \tau_i \leq t \leq T, \tag{14}$$

where

$$\mathbf{a}_{it} = \begin{cases} \mathbf{0} & t = \tau_i \\ \mathbf{a}_{i,t-1} + \boldsymbol{\eta}_{it} & t = \tau_i + 1, ..., T, \end{cases} \tag{15}$$

$\mathbf{0}$ denotes a matrix of zeros of appropriate dimension, and $\mathbf{v}_i, \boldsymbol{\eta}_{it}$ and $\mathbf{e}_{it}$ are 4-dimensional vectors that are assumed to have independent, multivariate normal distributions:

$$\mathbf{v}_i \sim \mathcal{IN}(\mathbf{0}, \boldsymbol{\Sigma}_v), \quad \boldsymbol{\eta}_{it} \sim \mathcal{IN}(\mathbf{0}, \boldsymbol{\Sigma}_\eta), \ \mathbf{e}_{it} \sim \mathcal{IN}(\mathbf{0}, \boldsymbol{\Sigma}_e). \tag{16}$$

We have an unbalanced panel data set, where firm i is observed from year $\tau_i \geq 1$ until $T_i \leq T$, where τ_i is the date of the firm's birth. The birth dates τ_i have an exogenous distribution, while the exit dates T_i can be endogenous.

When interpreting equation (14) in view of the structural equation (7), the term $\mathbf{a}_{it}$ corresponds to $\boldsymbol{\theta}_A \ln(A_{it}/A_{i1})$ and $\mathbf{v}_i$ corresponds to $\boldsymbol{\theta}_A \ln(A_{i1})$. On the other hand, in view of the extended model (12), $\mathbf{a}_{it}$ can be interpreted as $\boldsymbol{\theta}_A (\boldsymbol{\alpha}_{it} - \boldsymbol{\alpha}_{i1})$ and $\mathbf{v}_i$ as $\boldsymbol{\theta}_A \boldsymbol{\alpha}_{i1}$. All transient shocks and measurement errors are captured by $\mathbf{e}_{it}$, while all industry wide effects are captured by the intercept vector $\boldsymbol{\theta}_t$. It may seem restrictive to assume that $\mathbf{a}_{it}$ is a random walk, but our results are robust towards moderate departures from the random walk assumption; for example if the $\mathbf{a}_{it}$ process is slightly mean reverting, as suggested by Blundell and Bond (1999, 2000).

The structure of the covariance matrices are essential for the interpretation and identification of the model (14)-(16), which encompasses some well-known econometric models of firm heterogeneity as special cases: If $\boldsymbol{\Sigma}_\eta = \mathbf{0}$, we obtain the fixed effect model widely used to account for firm heterogeneity in the econometric panel data literature. When

$\boldsymbol{\Sigma}_e = \mathbf{0}$, the model is consistent with Gibrat's law discussed by Sutton (1997), where firm growth from period $t-1$ to t is independent of the level in period $t-1$. On the other hand, when $\boldsymbol{\Sigma}_e$ is a non-zero matrix, the model (14)-(16) implies "mean reversion", in the sense that any component of $\Delta \mathbf{y}_{it}$ will be negatively correlated with the corresponding component of $\mathbf{y}_{i,t-1}$[11].

A crucial point is whether the parameters of the covariance matrices are identified. Consider a sample covering two years; $t = 1, 2$. From (14)-(16), ignoring capital for simplicity, we have:

$$\text{Cov}\left(\mathbf{y}_{it}, \mathbf{y}_{is}\right) = \begin{cases} \boldsymbol{\Sigma}_v + \boldsymbol{\Sigma}_\eta \left[\min\left(t, s\right) - 1\right] & t \neq s \\ \boldsymbol{\Sigma}_v + \boldsymbol{\Sigma}_\eta (t - 1) + \boldsymbol{\Sigma}_e & t = s. \end{cases} \tag{17}$$

We then obtain: $\text{Cov}(\mathbf{y}_{i2}, \mathbf{y}_{i1}) = \boldsymbol{\Sigma}_v$, $\text{Cov}(\mathbf{y}_{i1}, \mathbf{y}_{i1}) = \boldsymbol{\Sigma}_v + \boldsymbol{\Sigma}_e$, and $\text{Cov}(\mathbf{y}_{i2}, \mathbf{y}_{i2}) = \boldsymbol{\Sigma}_v + \boldsymbol{\Sigma}_\eta + \boldsymbol{\Sigma}_e$.

Identification and testing of structural restrictions: As mentioned, there are no a priori constraints (apart from positive semi-definiteness) on the covariance matrices $\boldsymbol{\Sigma}_v$ and $\boldsymbol{\Sigma}_\eta$ in our general econometric model (14)-(16). We shall now consider the restrictions imposed by our structural models.

Let us first examine the single-factor model: According to (7)-(8), $\boldsymbol{\Sigma}_v$ and $\boldsymbol{\Sigma}_\eta$ can be factorized as:

$$\begin{aligned} \boldsymbol{\Sigma}_v &= \boldsymbol{\theta}_A \boldsymbol{\theta}_A{}' \, Var\left(\ln A_{i1}\right) \\ \boldsymbol{\Sigma}_\eta &= \boldsymbol{\theta}_A \boldsymbol{\theta}_A{}' \, Var\left[\ln\left(A_{it}/A_{i1}\right)\right], \end{aligned} \tag{18}$$

where we, for simplicity of notation, have assumed that $\tau_i = 1$. It should be noted that we cannot identify $\boldsymbol{\theta}_A$, since the variances in (18) are unknown.

If (18) holds, the rank of $\boldsymbol{\Sigma}_\eta$ is 1, and all components of the vector $\boldsymbol{\eta}_{it}$ are determined by a single latent factor, say z_{it}:

$$\boldsymbol{\eta}_{it} = \boldsymbol{\Gamma}_\eta z_{it}, \text{ with } z_{it} \sim \mathcal{IN}(0, 1), \tag{19}$$

where $\boldsymbol{\Gamma}_\eta$ is a 4×1 vector such that $\boldsymbol{\Sigma}_\eta = \boldsymbol{\Gamma}_\eta \boldsymbol{\Gamma}_\eta{}'$. From (15) and (19):

$$\mathbf{a}_{it} = \boldsymbol{\Gamma}_\eta \sum_{s=2}^{t} z_{is} \text{ for } t > 1. \tag{20}$$

---

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

Similarly, $\mathbf{v}_i$ can be expressed by a single latent factor z_{i1}:

$$\mathbf{v}_i = \boldsymbol{\Gamma}_v z_{i1}, \text{ with } z_{i1} \sim \mathcal{IN}(0,1), \tag{21}$$

where $\boldsymbol{\Gamma}_\eta$ is a 4×1 vector such that $\boldsymbol{\Sigma}_v = \boldsymbol{\Gamma}_v \boldsymbol{\Gamma}_v{}'$. From the definition of $\boldsymbol{\theta}_A$ in (16), a testable implication of this structural model is that the first three components within each vector $\boldsymbol{\Gamma}_\eta$ and $\boldsymbol{\Gamma}_v$ should be equal.

Preceding a test of the structure of $\boldsymbol{\Gamma}_\eta$ and $\boldsymbol{\Gamma}_v$, we must examine a more basic question: How well does a model with only one latent factor - i.e. where the rank of $\boldsymbol{\Sigma}_v$ and $\boldsymbol{\Sigma}_\eta$ is one - fit the data compared to a more general model with several latent factors? In particular, we would like to compare with the extended model (11), which implies that the rank of both $\boldsymbol{\Sigma}_v$ and $\boldsymbol{\Sigma}_\eta$ is equal to three.

First, consider a $\boldsymbol{\Sigma}_\eta$-matrix with general rank r. The innovations $\boldsymbol{\eta}_{it}$ can then be represented as:

$$\boldsymbol{\eta}_{it} = \boldsymbol{\Gamma}_\eta \mathbf{z}_{it}, \text{ with } \mathbf{z}_{it} \sim \mathcal{IN}(\mathbf{0}, \mathbf{I}_r), \tag{22}$$

where $\boldsymbol{\Gamma}_\eta = [\gamma_{ij}^\eta]_{4 \times r}$ is a $4 \times r$ matrix such that $\boldsymbol{\Sigma}_\eta = \boldsymbol{\Gamma}_\eta \boldsymbol{\Gamma}_\eta'$, $\mathbf{z}_{it}$ is an r-dimensional random vector and $\mathbf{I}_r$ is the identity matrix of order r. Similarly, we can express $\mathbf{v}_i$ as:

$$\mathbf{v}_i = \boldsymbol{\Gamma}_v \mathbf{z}_{i1}, \text{ with } \mathbf{z}_{i1} \sim \mathcal{IN}(\mathbf{0}, \mathbf{I}_r),$$

where $\boldsymbol{\Gamma}_v = [\gamma_{ij}^v]_{4 \times r}$ and $\boldsymbol{\Sigma}_v = \boldsymbol{\Gamma}_v \boldsymbol{\Gamma}_v{}'$. If the single-factor model is correct, $r = 1$, and we expect that the largest eigenvalue of the *estimated* covariance matrices $\widehat{\boldsymbol{\Sigma}}_\eta$ and $\widehat{\boldsymbol{\Sigma}}_v$ should be large relative to the others.

An interesting question, which we now will address, is whether we can identify the parameters $\boldsymbol{\theta}_A$ in the *extended* model (12) from $\boldsymbol{\Gamma}_\eta$: Let $\sigma_{A^*}^2$, σ_c^2, and σ_b^2 denote the variance of $\Delta \ln A_{it}^*$, $\Delta \ln c_{it}$, and $\Delta \ln b_{it}$, respectively, and define:

$$\mathbf{S} = \begin{bmatrix} \sigma_{A^*} & 0 & 0 \\ 0 & \sigma_c & 0 \\ 0 & 0 & \sigma_b \end{bmatrix}, \tag{23}$$

which – of course – cannot be identified. Let $\sigma_{cb} = \text{corr}(\Delta \ln c_{it}, \Delta \ln b_{it})$, with $\sigma_{cb} < 0$ (labor augmenting innovations reduce variable factor costs). Then, if $\boldsymbol{\Gamma}_\eta$ have the following structure:

$$\boldsymbol{\Gamma}_\eta = \begin{bmatrix} \gamma_1 & \gamma_2 & 0 \\ \gamma_1 & \gamma_3 & 0 \\ \gamma_1 & \gamma_4 & \gamma_5 \\ \gamma_6 & \gamma_7 & 0 \end{bmatrix} \tag{24}$$

14

with $\gamma_1 > 0$, $\gamma_2 < 0$ and where $\Delta \ln A_{it}^*$ and $(\Delta \ln c_{it}, \Delta \ln b_{it})$ are mutually independent, the following holds:

$$
\begin{aligned}
\boldsymbol{\Gamma}_\eta \mathbf{z}_{it} &= \boldsymbol{\theta}_A \Delta \boldsymbol{\alpha}_{it} \\
\mathbf{D} \mathbf{z}_{it} &= \mathbf{S}^{-1} \Delta \boldsymbol{\alpha}_{it} \\
\boldsymbol{\Gamma}_\eta \mathbf{D}^{-1} &= \boldsymbol{\theta}_A \mathbf{S},
\end{aligned}
\tag{25}
$$

where

$$
\mathbf{D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \sigma_{cb} & \sqrt{1 - \sigma_{cb}^2} \end{bmatrix}
\tag{26}
$$

is uniquely identified from $\boldsymbol{\Gamma}_\eta$. Equations (25)-(26) say that we can identify the *standardized* innovations $\mathbf{S}^{-1} \Delta \boldsymbol{\alpha}_{it}$ from $\mathbf{z}_{it}$ (any non-zero mean of $\Delta \boldsymbol{\alpha}_{it}$ is absorbed into $\boldsymbol{\theta}_t$) and the structural coefficients $\boldsymbol{\theta}_A$ from $\boldsymbol{\Gamma}_\eta$ – except for postmultiplication with the unknown diagonal matrix $\mathbf{S}$. That is, each column of $\boldsymbol{\theta}_A$ is identified up to an unknown scaling factor. A proof of (25)-(26) is found in Appendix B.

A general 4×4 covariance matrix $\boldsymbol{\Sigma}_\eta$ with rank 3 has a minimal representation: $\boldsymbol{\Sigma}_\eta = \boldsymbol{\Gamma}_\eta \boldsymbol{\Gamma}_\eta{}'$ where $\boldsymbol{\Gamma}_\eta$ is a 4×3 matrix with zeros above the diagonal, and therefore contains 9 free parameters. On the other hand, the specification of $\boldsymbol{\Gamma}_\eta$ in (24) contains 7 free parameters. Thus there are two overidentifying restrictions embedded in (25)-(26), which we will subject to statistical testing in Section 7.

Attrition: A question that we need to address is how we should account for sample attrition? Since attrition may be caused by an exit decision that depends on a firm's efficiency, this problem is potentially important in our case.

We propose a simultaneous equation model which is an extension of the model due to Heckman (1978): Suppose that u_{it} is a latent variable related to the net value of the firm such that when $u_{it} > 0$ the firm will continue to operate, while when $u_{it} < 0$ it is decided to close down the firm. Hence, we define an indicator variable, χ_{it}, such that $\chi_{it} = 1$ if $u_{it} > 0$ and $\chi_{it} = -1$ if $u_{it} < 0$.

Our reduced form model of attrition has the following form:

$$
\begin{aligned}
u_{it} &= \boldsymbol{\gamma}_A'(\mathbf{v}_i + \mathbf{a}_{i,t+1}) + \boldsymbol{\gamma}_x' \mathbf{x}_{it} + \gamma_t + \widetilde{\varepsilon}_{it} \\
\chi_{it} &= \text{sign}(u_{it})
\end{aligned}
\qquad t = 1, ..., T_i,
\tag{27}
$$

15

where $\boldsymbol{\gamma}_A$, $\boldsymbol{\gamma}_x$, γ_t are unknown parameters, $\mathbf{x}_{it}$ is a vector of observed variables up until t, and $\widetilde{\varepsilon}_{it}$ is normally distributed. Note that under the structural single-factor model (7), $\boldsymbol{\gamma}'_A(\mathbf{v}_i + \mathbf{a}_{i,t+1})$ is proportional to $\ln A_{i,t+1}$ for any vector $\boldsymbol{\gamma}_A$ (which therefore is not identified). In this case, (27) says that a firm will exit at the beginning of year $t+1$ if its efficiency, $A_{i,t+1}$, falls short of a threshold depending on observed variables, $\mathbf{x}_{it}$. A similar exit rule was derived by Olley and Pakes (1996). In their model, $\boldsymbol{\gamma}'_x\mathbf{x}_{it}$ is a polynomial in investment and capital, meant to approximate a general, unknown function.

The classical Marshallian exit rule provides a motivation for our model specification, and some guidance with respect to which variables to include in $\mathbf{x}_{it}$. According to Marshall (1966, p. 349), "production is likely to come to a sharp stop" when "the price falls so low that it does not pay for the out of pocket expenses." An interpretation of this statement is that the probability of firm exit increases as the short-run profit function, $\Pi(A_{i,t+1}, K_{it}, P_{t+1})$, decreases relative to the capital costs:

$$\Pi(A_{i,t+1}, K_{it}, P_{t+1}) < e^{-\widetilde{\varepsilon}_{it}}(r+\delta)q_{K,t+1}K_{it}, \tag{28}$$

where $\widetilde{\varepsilon}_{it}$ is the error term, r is the interest rate and δ is the depreciation rate of capital. If the profit function is homogeneous in $A_{i,t+1}$ and K_{it} with diminishing returns to scale, we obtain, after taking logarithms of both sides of (28), an equation of the form (27) with $\mathbf{x}_{it} = \ln K_{it}$.

The equations (27) should be considered as a reduced form expression; a statistical model which enables us to predict attrition, not to explain it. Thus we are in line with Abowd et al. (2001), who propose a flexible statistical approach to the sample self-selection problem which do not rely on formal economic modelling of the decisions which cause attrition. In accordance with their line of argument, we can think of (27) as an auxiliary model whose parameters are not of interest per se, but which enables us to control for self-selection bias when making inferences about the structural model. However, in contrast to Abowd et al. (2001) our model of attrition does not rely on the missing at random (MAR) assumption,[12] and therefore entails some demanding computational challenges which we address in Appendix A.

---

[12] The missing at random (MAR) condition says that information about whether the firm is operative in year $t+1$ should not help us to predict $\mathbf{y}_{i,t+1}$, given the *history* of the observed variables: $\mathbf{y}_{i1}, ..., \mathbf{y}_{it}$. See Little and Rubin (1987); and Moffitt, Fitzgerald and Gottschalk (1999). The latter authors refer to the MAR condition as "selection on observables".

The likelihood function: We will here outline how we can arrive at a convenient expression for the likelihood function. Estimation issues are considered in detail in Appendix A.

We first note that if $\tau_i = 1$, we can substitute for $\mathbf{v}_i + \mathbf{a}_{i,t+1}$ in (27), using (14) and the fact that $\mathbf{a}_{i,t+1} = \mathbf{a}_{it} + \mathbf{\Gamma}_\eta \mathbf{z}_{i,t+1}$, to obtain the following representation of our model:

$$
\begin{aligned}
\mathbf{y}_{i1} &= \boldsymbol{\theta}_K \ln K_{i,0} + \boldsymbol{\theta}_1 + \mathbf{\Gamma}_v \mathbf{z}_{i1} + \mathbf{e}_{i1} \\
\Delta \mathbf{y}_{it} &= \boldsymbol{\theta}_K \Delta \ln K_{i,t-1} + \Delta \boldsymbol{\theta}_t + \mathbf{\Gamma}_\eta \mathbf{z}_{it} + \mathbf{e}_{it} - \mathbf{e}_{i,t-1}, \ t = 2, 3, ... \\
u_{it} &= \boldsymbol{\gamma}'_A \mathbf{y}_{it} + \boldsymbol{\gamma}'_x \mathbf{x}_{it} + \gamma_t + \varepsilon_{it}, \ t = 1, 2, ... \\
\chi_{it} &= \text{sign}(u_{it}), \ t = 1, 2, ...
\end{aligned}
\tag{29}
$$

with

$$
\varepsilon_{it} = \widetilde{\varepsilon}_{it} + \boldsymbol{\gamma}'_A (\mathbf{\Gamma}_\eta \mathbf{z}_{i,t+1} - \mathbf{e}_{it}),
\tag{30}
$$

and $\text{Var}(\varepsilon_{it}) = 1$ by normalization. Furthermore, $\boldsymbol{\gamma}_x$, $\mathbf{x}_{it}$, and γ_t have been redefined so as to be consistent with the transformation from (27) and (14) into (29); in particular, $\mathbf{x}_{it}$ is augmented with $\ln K_{i,t-1}$. We see from (30) that ε_{it} may be correlated with *both* $\mathbf{e}_{it}$ and $\mathbf{z}_{i,t+1}$:

$$
\begin{pmatrix} \mathbf{e}_{it} \\ \mathbf{z}_{i,t+1} \\ \varepsilon_{it} \end{pmatrix} \sim \mathcal{IN} \left(\mathbf{0}, \begin{bmatrix} \mathbf{\Sigma}_e & \mathbf{0} & \boldsymbol{\lambda} \\ \mathbf{0} & \mathbf{I} & \boldsymbol{\rho} \\ \boldsymbol{\lambda}' & \boldsymbol{\rho}' & 1 \end{bmatrix} \right).
$$

In Appendix A it is shown that we can utilize properties of the multivariate normal distribution to obtain an explicit expression for the joint density of $\{\mathbf{y}_{i1}, ..., \mathbf{y}_{iT_i}, \chi_{i,T_i}\}$, although the maximization of the likelihood function requires simulation methods. An interesting special case is obtained when $\boldsymbol{\lambda} = \mathbf{0}$ and $\boldsymbol{\rho} = \mathbf{0}$. Then the MAR condition holds and a relatively simple closed form (partial) likelihood function can be derived. In our experience, maximization of the partial likelihood function (under the MAR assumption) provides excellent starting values when estimating the unrestricted model, even when $\boldsymbol{\lambda}$ or $\boldsymbol{\rho}$ are significantly different from zero.

5 Why do firms differ in efficiency?

Our econometric framework allows us to decompose differences in efficiency and to *quantify* the relative importance of permanent differences and cumulated innovations. In view of

our econometric model (14), a natural measure of the importance of permanent differences relative to idiosyncratic innovations in a given cohort of firms, say with age T, is:

$$V_T \equiv \frac{\text{tr } Var(\mathbf{v}_i)}{\text{tr } Var(\mathbf{a}_{iT})} = \frac{\text{tr } \boldsymbol{\Sigma}_v}{(T-1)\text{tr } \boldsymbol{\Sigma}_\eta}. \tag{31}$$

In the single-factor model (7), V_T reduces to

$$V_T = \frac{Var(\ln A_{i1})}{Var(\ln(A_{iT}/A_{i1}))} = \frac{\sigma_v^2}{(T-1)\sigma_\eta^2},$$

where σ_v^2 and σ_η^2 are the (non-zero) eigenvalues of $\boldsymbol{\Sigma}_v$ and $\boldsymbol{\Sigma}_\eta$, respectively.

The measure V_T defined in (31) ignores endogenous exit, which will tend to reduce the variance both in $\mathbf{v}_i$ and $\mathbf{a}_{iT}$. Hence, we focus on a modified version of (31): Let M_T denote the set of firms born in year 1 that are still operative in year T. We define the *conditional variance* ratio, CV_T, as

$$CV_T \equiv \frac{\text{tr } Var(\mathbf{v}_i | i \in M_T)}{\text{tr } Var(\mathbf{a}_{iT} | i \in M_T)}. \tag{32}$$

Thus, while V_T is computed from the *unconditional* distribution of the latent variables, CV_T is calculated from their conditional distribution given survival. Of course, it is impossible to obtain a closed form expression for CV_T, but it is straightforward to simulate data from our joint model of sales, factor demand and attrition, and thus generate random numbers from the distribution of $\mathbf{v}_i$ and $\mathbf{a}_{iT}$ conditional on survival. The conditional distribution depends on the initial distribution of capital, K_{i1}. Analogous to what is done in bootstrapping, we estimate the initial distribution of capital using the realized values of K_{i1} in each industry, from which we make random draws with replacement.

The extended model presented in Section 3.2 enables us to evaluate how much of the variation in the innovation vector $\boldsymbol{\eta}_{it}$ that is due to Hicks-neutral efficiency shocks. We propose the following relative variance measure, RV:

$$RV \equiv \frac{\text{tr } Var\left(\theta_{A,1}\Delta \ln(A_{it}^*)\right)}{\text{tr } \left(\boldsymbol{\Sigma}_\eta\right)},$$

where $\boldsymbol{\theta}_{A,1}$ is the first column of $\boldsymbol{\theta}_A$, i.e. the loading vector of $\ln(A_{it}^*)$ in (12). Since $\Delta \ln(A_{it}^*)$ is orthogonal to the other latent factors, $1 - RV$ can be interpreted as the relative variance of $\Delta \ln c_{it}$ and $\Delta \ln b_{it}$, when combined into a single residual factor.

6   Data and variable construction

We use a recently established database from Statistics Norway: the Capital database, which contains annual observations on fixed capital (tangible fixed assets), costs of rented capital (i.e. operational leasing), sales, wage costs, intermediates, man hours, and many other variables for all Norwegian joint stock companies in the manufacturing sector for the period 1993-2002[13]. The main statistical unit in the database is the firm: A firm is defined as "the smallest legal unit comprising all economic activities engaged in by one and the same owner" and corresponds in general to the concept of a company. A firm may consist of one or more establishments. The establishment is the geographically local unit doing economic activity within an industry class. The population of joint stock companies comprises about 80% of total manufacturing employment in 2002 (but a much smaller share of the total number of firms). The stock companies' employment-weighted share of the population of *new* firms is roughly the same. In this paper we analyze four relatively high-tech sectors: Rubber and plastic products (NACE 25); Machinery and equipment (NACE 29); Electrical and optical equipment (NACE 30-33), and Transport equipment (NACE 34-35). See Appendix C for a listing of the NACE sector codes.

The database combines information from mainly two sources: Accounts statistics for all Norwegian joint-stock companies, and Structural statistics for the manufacturing sector. Many of the variables in the database have been extensively revised and crosschecked against different data sources by Statistics Norway, including tax return forms. A very important feature of the database, is that it contains measurements of net capital stocks in both current and fixed prices.

The method for calculating the capital stocks in current prices is based on combining book values from the financial accounts with gross investment data.[14] Since our econometric model contains a single aggregate capital variable, we have constructed this as being proportional to the sum of the user cost of capital owned by the firm[15] and the

---

[13]See "Documentation of the capital database. A database with data for tangible fixed assets and other economic data at the firm level," which can be downloaded from: http://www.ssb.no/english/subjects/10/90/doc_200416_en/doc_200416_en.pdf

[14]See Raknerud, Rønningen and Skjerpen (2003) for technical details and a thorough evaluation of the data quality.

[15]Capital is divided into two groups of assets in the database: (i) Buildings and land (which have long service lives) and (ii) Other tangible assets (with small or medium service lives). Separate user cost estimates have been calculated for the two groups.

total operational leasing costs. This is consistent with using a (constant returns to scale) Cobb-Douglas aggregation function.

Our model contains four variables, which are measured on log-scale: sales, labor costs, materials, and capital. Labor costs incorporate salaries and wages in cash and kind, social security and other costs incurred by the employer. In general, all costs and revenues are measured in nominal prices, and incorporate direct taxes and subsidies, except VAT. We have not deflated the variables with available (industry wide) deflators as the econometric model contains an industry wide time-varying intercept vector.

Following Caves' (1998) survey of empirical findings on firm growth and turnover, we have not stressed the distinction between a firm and an establishment.[16] The unit of observation in our data set is the firm. About 10-20 percent of the establishments belong to multi-establishment firms in the sectors we consider.

Sometimes a firm may vanish from the database even if some of its establishments are still operating. Our data indicate whether the disappearance of a firm is due to (i) a close down of all production units or (ii) a merger, acquisition, or some other change of ownership structure. Only (i) is counted as a firm exit as defined in Section 4, while (ii) is considered as exogenous attrition.

Initially *all* firms in a sector that were operating during 1993-2002 were included in the sample. For firms established before 1993, we introduced separate (nuisance) parameters for the distribution of $\mathbf{v}_i$[17], since $\mathbf{v}_i$ for these firms is composed of both permanent differences and cumulated innovations (up until 1993) and therefore has a different meaning than for firms established during 1993-2002. The focus of the analysis of firm heterogeneity is on firms established during the observation period. However, sometimes a firm is registered as a new entrant to the industry although its establishments are old; for example if a large firm is split into several smaller firms. Only a firm that consists of new establishments (typically a single establishment firm) is considered as an entrant in this study.

Some "cleaning" of the data was performed. A firm was excluded from the sample if:

---

[16]Caves (1998) points out that most of the results on firm growth and turnover have been insensitive to the establishment-firm distinction. This is not to deny that the distinction between firms (or lines-of-business) and establishments raises interesting questions for our analysis. For instance, are there strong correlations between efficiency levels across establishments within a firm? Do new establishments from an existing firm have the same efficiency as new firms?

[17]That is, $\mathbf{v}_i \sim \mathcal{N}(\boldsymbol{\mu}_v, \widetilde{\Sigma}_v)$

(i) the value of an endogenous variable is missing for two or more subsequent years; (ii) the firm disappears from the raw data file and then reappears more than one time; or (iii) the firm is observed in a single year only. These trimming procedures reduced the data set by 10-15 percent. Some summary statistics are presented in Table 1.

7   Empirical results

This section, which presents our empirical results, is divided into three parts. First, we examine the empirical validity of the structural models presented in Section 3. We find that the single-factor model accounts quite well for the empirical patterns in all of the industries. Nevertheless, allowance for labor augmenting innovations is needed in order to explain the empirical autocorrelation patterns of labor productivity. We then show that permanent differences dominate differences generated by cumulated, firm-specific innovations in explaining observed firm heterogeneity. We also find that Hicks-neutral innovations dominate labor augmenting innovations in explaining firm growth. Finally, we examine the nature of sample self-selection.

7.1   The validity of the structural models

Table 2 presents the estimated eigenvalues from the factor decompositions described in Section 4. The second column presents the estimated eigenvalues of the covariance matrix for the idiosyncratic innovations, $\boldsymbol{\Sigma}_\eta$, when no rank restrictions are imposed a priori. In all the industries, the largest eigenvalue is at least an order of magnitude larger than the second largest eigenvalue. The same pattern is present in the third column, presenting the estimated eigenvalues of the covariance matrix of the permanent differences, $\boldsymbol{\Sigma}_v$.

These patterns of eigenvalues show that the persistent differences in performance can largely be summarized by one latent factor determining all components of $\boldsymbol{\eta}_{it}$ and $\mathbf{v}_i$, respectively. This conclusion is confirmed by the last columns of Tables 2 and 3, which presents (pseudo-) R^2-measures varying between .88 and .91 when no rank restrictions are imposed on $\boldsymbol{\Sigma}_\eta$ and $\boldsymbol{\Sigma}_v$ (Table 2), and between .86 and .89 in the single-factor model (Table 3)[18]. Thus, there is only a small increase in R^2 when going from the rank-one to

---

[18]Our pseudo R^2-measure is

$$R^2 = 1 - \frac{\operatorname{tr} \widehat{Var}(\widehat{\mathbf{e}}_{it})}{\operatorname{tr} \widehat{Var}(\mathbf{y}_{it} - \widehat{\boldsymbol{\theta}}_t)},$$

the rank-four model.

The fourth column in Table 2 depicts the eigenvalues of $\boldsymbol{\Sigma}_e$, the covariance matrix associated with transient shocks. The results show that the transient shocks are not dominated by a single, common latent factor, in contrast to the persistent shocks. That is, transient fluctuations are not common across the four endogenous variables. We notice that the variance generated by the transient variance component is larger than the variance of the innovation component in all industries: $\mathrm{tr}(\boldsymbol{\Sigma}_e) > \mathrm{tr}(\boldsymbol{\Sigma}_\eta)$. The transient fluctuations account for considerable mean reversion in the dynamic process for the observable variables, as pointed out in Section 4.

The single-factor model: The structural model presented in Section 3.1 does not only impose a rank condition on $\boldsymbol{\Sigma}_\eta$ and $\boldsymbol{\Sigma}_v$. These matrices should also have the structure that follows from $\boldsymbol{\theta}_A$; see Section 3.1 and, in particular, (8) and (18). That is, the single-factor model in Section 3.1 requires that the three first components within each loading vector $\boldsymbol{\Gamma}_\eta$ and $\boldsymbol{\Gamma}_v$ should be the same. The estimates for the factor loadings $\boldsymbol{\Gamma}_\eta$ and $\boldsymbol{\Gamma}_v$ in the single-factor model are presented in Table 3, with standard errors in parentheses.

Formal χ^2-tests of the structural restrictions on $\boldsymbol{\Gamma}_\eta$ and $\boldsymbol{\Gamma}_v$ and on the capital coefficients, $\boldsymbol{\theta}_K$, are presented in Table 4. Except for a weak rejection of the restrictions on $\boldsymbol{\Gamma}_v$ in the sector Electrical and optical equipment (NACE 30-33) and on the capital coefficients, $\boldsymbol{\theta}_K$, in the sector Machinery (NACE 29), the structural hypotheses are maintained for the other three sectors. Considering the relatively large number of tests reported in Table 4, the overall results can be seen as generally supportive of the one-factor model.

Nevertheless, not all aspects of the data are explained well by this simple model. First, the second largest eigenvalues of $\boldsymbol{\Sigma}_\eta$ and $\boldsymbol{\Sigma}_v$, albeit small, are clearly significant in all sectors in view of the small standard errors. Thus, the hypothesis of a single latent factor is rejected. One might be tempted to dismiss this conclusion, since it is well known that the rejection of any null-hypothesis is only a question of having a sufficiently large data set[19]. On the other hand, it was noted already in Section 2 that Hicks-neutral efficiency shocks cannot explain the autocorrelation pattern of labor productivity depicted in Figure

---

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

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

22

2.

Figure 4 elaborates on this point: The three lowest graphs in the figure show the autocorrelation function of log labor productivity (conditional on survival) for: (i) the actual data; (ii) the estimated single factor model (with no restrictions on the factor loadings); and (iii) the estimated extended model (11) with *three* latent factors and with the overidentifying restrictions (24) imposed on $\mathbf{\Gamma}_\eta$.

We see that the single-factor model systematically underpredicts the empirical auto-correlations, even if we have imposed no equality restrictions on its loading coefficients. Imposing such restrictions would make the situation even worse, as all autocorrelations would become zero. On the other hand, the extended model fits the autocorrelations in labor productivity remarkably well – with a possible exception at the first lag where the empirical autocorrelation coefficient is somewhat higher than predicted by the extended model.

The upper three graphs in Figure 4 show the corresponding autocorrelation functions for log sales. We first note that both econometric models fit the empirical autocorrelation well. This may explain why the R^2- measures reported in Table 2 and Table 3 are so similar. Our assumption that persistent differences in firm performance evolves as a random walk, in accordance with Gibrat's law[20], seems to be substantiated by the graphs in Figure 4.

Let us now turn to the capital coefficient estimates. First, we note that the loading coefficients of the latent variables in the capital accumulation equation reported in columns 2 and 3 of Table 3 is about 1/3 of the loading coefficients in the sales and factor demand equations. This indicates that an innovation which increases sales and factor demand with 1 percent, increases the capital stock with around 0.3 percent. This is a significant effect, suggesting a clear link between innovations and investments.

The coefficients of lagged capital, $\ln K_{i,t-1}$, in each of the four equations in the system (7) are presented in the fourth column in Table 3. The coefficients are around .7 in the fourth (capital) equation in most sectors, which show that the speed of the adjustment of capital towards its equilibrium is moderate. Price-taking behavior and constant returns

---

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

to scale imply that the coefficient of $\ln K_{i,t-1}$ in the sales and factor demand equations, $\gamma/(1-\varepsilon)$, should be equal to 1, while our estimates lie between .05 and .17. Thus, we find clear evidence of diminishing returns to scale.

Even though we cannot identify the parameters γ and ε of the production function (1), it is interesting to use the "budget shares" $\widehat{\varepsilon} = \sum_i (w_t^l L_{it} + w_t^m M_{it})/\sum_i S_{it}$ and $\widehat{\gamma} = \sum_i (\delta + r)q_t K_{i,t-1}/\sum_i S_{it}$ as benchmarks, since these are widely used as estimators in practice. We find that the budget shares obtained by lumping data of all the sectors together are stable over time, with $\widehat{\varepsilon} \approx .9$ and $\widehat{\gamma} \approx .05$, suggesting that $\gamma/(1-\varepsilon) \approx .5$. The latter estimate seems too high to be consistent with our results. Therefore, the assumption of price-taking behavior may not be plausible. It is interesting to note that in the monopolistic competition model (11), the coefficient of capital is: $\gamma (e-1)/(\varepsilon + e - e\varepsilon)$, which is small when the price elasticity of demand, e, is large. For example: $\varepsilon = .9$, $\gamma = .05$, and $e = 3.5$, gives a coefficient of capital equal to .1, which is in the middle of our range of capital coefficient estimates. Earlier studies on Norwegian manufacturing data also provide evidence of imperfect competition and market power in many industries (see Klette, 1999).

The extended model: The extended model (11) contains three latent factors: the Hicks-neutral efficiency parameter, $\ln A_{it}^*$ (which also incorporates demand idiosyncracies due to quality differences between different producers), the variable factor price index, $\ln c_{it}$, and the labor augmenting parameter, $\ln b_{it}$. It was noted in Section 3 that labor productivity will depend on $\ln c_{it}$ and $\ln b_{it}$, but not on $\ln A_{it}^*$. Based on the identifying restrictions discussed in Section 4 and Appendix B, we are able to identify the structural coefficient matrix $\boldsymbol{\theta}_A$ up to an unknown scale coefficient for each column. That is, we are able to identify the following parameters:

$$
\begin{aligned}
\widetilde{\theta}_1 &= \frac{(e-1)}{\varepsilon + e - e\varepsilon}\sigma_A, \ \widetilde{\theta}_2 = -\frac{\varepsilon(e-1)}{\varepsilon + e - e\varepsilon}\sigma_c, \ \widetilde{\theta}_3 = \left(-\frac{\varepsilon(e-1)}{\varepsilon + e - e\varepsilon} + \frac{\rho}{1-\rho}\right)\sigma_c \\
\widetilde{\theta}_4 &= \frac{\rho}{1-\rho}\sigma_b, \ \widetilde{\theta}_5 = (1-\phi)\kappa_A\sigma_A, \ \widetilde{\theta}_6 = (1-\phi)\kappa_c\sigma_c, \\
\sigma_{cb} &= \text{Corr}(\ln c_{it}, \ln b_{it}).
\end{aligned}
$$

The estimates are depicted in Table 5. The coefficients of $\ln K_{i,t-1}$ are identical to the estimates in Table 3 up to two decimal places, and are therefore not reported.

24

First, we note that in all sectors the estimated loading coefficient of $\ln A_{it}^*$ in the first three equations, $\widetilde{\theta}_1$, is a weighted average of the three first components of $\mathbf{\Gamma}_\eta$ reported in Table 3. Moreover, the estimated coefficient of $\ln A_{it}^*$ in the capital equation, $\widetilde{\theta}_5$, is very close to the fourth component of $\mathbf{\Gamma}_\eta$ in all sectors. With respect to the other latent factors, our results indicate that $\widetilde{\theta}_4 < 0$ in all sectors, except in Plastics (NACE 25). Thus, labor augmenting innovations are mostly labor saving, since $\widetilde{\theta}_4 < 0$ is equivalent with $\rho < 0$. Note that $\widetilde{\theta}_4 < 0$ is equivalent with $\widetilde{\theta}_3 < \widetilde{\theta}_2$, which holds also for the estimated parameters. The estimates of σ_{cb} in Table 5 shows that the labor augmenting parameter, $\ln b_{it}$, and the variable factor price index, $\ln c_{it}$, are highly negatively correlated, as we expected. Finally, the last column of Table 5 reports the result of a χ^2-test (with 2 degrees of freedom) of the overidentifying restrictions embedded in the extended model. The overidentifying restrictrions are maintained in all sectors.

7.2 Permanent differences dominate

We can now examine the origin and evolution of differences in performance across firms. Table 6 presents measures of the magnitude of permanent differences and differences generated by cumulated innovations within each of the four industries based on the estimated extended model. Column 2 presents the ratio of the trace of the variance-covariance matrix of the permanent differences, $\text{tr}(\mathbf{\Sigma}_v)$, to the trace of the variance-covariance matrix of the cumulated innovations, $\text{tr}(\mathbf{\Sigma}_\eta)$. This ratio shows how many years innovations must be accumulated in order to account for as much of the heterogeneity as the permanent differences. These figures lie between 12 and 31, showing that permanent differences dominate over idiosyncratic innovations in the cohort of firms established in 1993 in all sectors.

These results do not, however, provide a fully satisfactory measure of the importance of permanent differences in explaining the observed variation in firm performance, since they neglect the issue of exit and self-selection. We argued in Section 5 that a better measure is provided by the *conditional* variance ratio, CV_T. The conditional variance ratio among firms surviving from 1993 until (at least) 2002 is presented in column 4 for each industry. The conditional variance ratios vary from 1.6 in Machinery (NACE 29) to 3.0 in Electrical and optical equipment (NACE 30-33). The pattern from the previous columns remains, i.e. the variance of the permanent differences is larger than

the variance in the cumulated, idiosyncratic innovations in all industries. We also find that the conditional variance ratios are somewhat lower than the corresponding unconditional variance ratios presented in column 3.

Finally, as explained in Section 5, the variance ratio RV measures the relative importance of Hicks-neutral innovations, i.e. shocks in $\ln A_{it}^*$, for firm growth. The results in the last column of Table 6, show that Hicks-neutral efficiency shocks account for between 70-80 percent of the variance in cumulated innovations. This explains why labor productivity, which is correlated with $\Delta \ln b_{it}$ and $\Delta \ln c_{it}$, but not $\Delta \ln A_{it}^*$, are only weakly related to firm growth, as depicted in Figure 3.

7.3   Sample self-selection

Table 7 presents likelihood ratio tests of the hypotheses that $\boldsymbol{\rho} = \mathbf{0}$ and $\boldsymbol{\lambda} = \mathbf{0}$ for the model with one latent factor in $\boldsymbol{\eta}_{it}$ and $\mathbf{v}_i$. The estimated correlation between the innovation, $\boldsymbol{\eta}_{i,t+1}$, and firm exit in year t is negative. A likelihood ratio test of the restriction that this correlation is zero: $\boldsymbol{\rho} = \mathbf{0}$, is presented in the second column in Table 7 and is rejected at the 5 percent level of significance in all industries. Thus there is significant selection systematically eliminating firms with low efficiency[21]. On the other hand, the results in the third column of Table 7 show that there is no connection between the transitory error term, $\mathbf{e}_{it}$, and firm exit (conditional on the observed variables $\mathbf{x}_{it}$).

The values of the χ^2 statistics in Table 7 change little when going from the single-factor model, to the extended model. We find that self-selection is significantly related to Hicks-neutral innovations, but not to the other latent factors.

In all industries we find a *negative* correlation between the permanent efficiency levels, $\mathbf{v}_i$, and the subsequent innovations, $\mathbf{a}_{iT}$, among surviving firms. Our interpretation of this negative correlation is that a firm with a low permanent efficiency level must have a high growth in efficiency in its subsequent years in order to survive and *vice versa*. That is to say, selection is based on the firm's overall efficiency, which is the combination of the permanent efficiency levels and the innovations.

---

[21]Similar findings have been presented in a number of studies, as surveyed by Foster, Haltiwanger and Krizan (2001). However, our measurement of efficiency differs from the previous studies. The negative correlation between the probability of exit and a firm's productivity level has not been striking in previous studies of Norwegian manufacturing firms, see Møen (1998).

8 Conclusion

This paper examines the large differences across firms in terms of sales and demand for labor, materials and capital. With firm level observations from four manufacturing industries covering 10 years, we showed that most of the differences in sales and factor demands can be accounted for by a structural model with price-taking behavior, fully optimizing supply and factor demand, and a simple production function with Hicks-neutral efficiency shocks.

Nevertheless, the relatively low estimated output and input elasticities of capital, indicate that the assumption of price taking behavior is too simple. A more satisfactory model is one of monopolistic competition, allowing differences both in efficiency and product characteristics – e.g. quality differences – across different producers within the sector. Furthermore, in order too account for persistent differences in labor productivity across firms, we are led to include a labor-augmenting latent factor in the production function. We find that Hicks-neutral innovations dominate labor augmenting innovations in explaining firm growth.

The structural model enables us to investigate the origin and evolution of the differences in performance across firms. The empirical results show that *permanent* differences dominate among the firms established within the 10-year period we consider, as they exceed differences in cumulated innovations by a factor ranging between about 1.5 and 3. across the four high-tech industries.

The most striking and controversial result from our analysis is its implications for efficiency measurement. We argue that size is a better indicator of efficiency than labor productivity, as long as we also account for the fixity of capital. It is well known that differences in firm size should reflect differences in efficiency, while the serious problem we point out with labor productivity as a measure of efficiency differences seems to have been largely neglected in the literature[22].

---

[22]See, however, Bernard et al. (2000) and Klette and Kortum (2004).

References

Abel,A. and J. Eberly (1996): Optimal investment with costly reversibility. *Review of Economic Studies*, 63,581–93.

Abowd, J., F. Kramarz and D. Margolis (1999): High wage workers and high wage firms. *Econometrica*, 67, 251–333

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

Aghion, P. and P. Howitt (1994): Growth and unemployment. *Review of Economic Studies*, 61,477–94.

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

Baily,M, C. Hulten and D. Campbell (1992): Productivity Dynamics in Manufacturing Plants. *Brookings Papers on Economic Activity. Microeconomics*, 187-249.

Baumol, W. and E. Wolff (1984): On interindustry differences in absolute productivity. *Journal of Political Economy* ,92, 1017-34.

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

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

Binswanger, H. (1974): The measurement of technical change biases with many factors of production. *The American Economic Review*, 64, 964–76.

Bloom, N. (2000): The real options effect of uncertainty on investment and labour demand. Institute for Fiscal Studies, IFS Working Papers: W00/15.

Bloom,N., S. Bond and J. Van Reenen (2001): The dynamics of investment under uncertainty. Institute for Fiscal Studies, IFS Working Papers: W01/05.

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

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

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

Caballero, R., E. Engel, and J. Haltiwanger (1995): Plant-level adjustment and aggregate investment dynamics. Brookings Papers on Economic Activity, 1–39.

— (1999): Explaining investment dynamics in U.S. manufacturing: A generalized (S, s) approach. *Econometrica*, 67, 783–826.

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

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

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

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

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

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

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.

— (2002): The link between aggregate and micro productivity growth: Evidence from retail trade. *US Census, Economic Studies 2002.*

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

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

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

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

Hajivassiliou, V., D. McFadden, and P. Ruud (1996): Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results. *Journal of Econometrics*, 72, 85–134.

Heckman, J. (1978): Dummy endogeneous variables in a simulataneous equation system. *Econometrica*, 46, 931–59.

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

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

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

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

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

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

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

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

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

Klette, T. (1999): Market power, scale economics and productivity: Estimates from a panel of establishment data. *The Journal of Industrial Economics*, 110, 451–75.

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 (2004): Innovating firms and aggregate innovation. *Journal of Political Economy*, 112, ?–?.

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

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

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

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

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

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

Lutkepohl, H. (1996): *Principles of Economics.*. 8th edition. London: Macmillan.

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.

Nilsen,O.A. and F. Schiantarelli (2003): Zeros and lumps in investment: Empirical evidence on irreversibilities and nonconvexities. *Review of Economics and Statistics*, 85,1021–37.

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.

Sargent, T. (1978): Estimation of Dynamic Labor Demand Schedules under Rational Expectations. *Journal of Political Economy*, 86, 1009–44.

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

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

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

Appendix A: Estimation

In order to derive the full likelihood function, we re-write the system (29) using properties of multivariate regression:

$$
\begin{aligned}
\mathbf{z}_{i,t+1} &= \boldsymbol{\rho}\varepsilon_{it} + M\widetilde{\mathbf{z}}_{i,t+1} \\
\mathbf{e}_{it} &= \boldsymbol{\lambda}\varepsilon_{it} - \boldsymbol{\lambda}(M^{-1}\boldsymbol{\rho})'\widetilde{\mathbf{z}}_{i,t+1} + \widetilde{\mathbf{e}}_{it}
\end{aligned}
$$

where $\widetilde{\mathbf{z}}_{i,t+1} \sim \mathcal{IN}(\mathbf{0},\mathbf{I})$, M is lower triangular, $\widetilde{\mathbf{e}}_{it} \sim IN(0,\widetilde{\Sigma}_e)$ and ε_{it}, $\widetilde{\mathbf{z}}_{it}$ and $\widetilde{\mathbf{e}}_{it}$ are white noise and independent of each other. Defining $\widetilde{\boldsymbol{\Gamma}}_\eta = \boldsymbol{\Gamma}_\eta M$ and $\widetilde{\boldsymbol{\rho}} = M^{-1}\boldsymbol{\rho}$, we can write:

$$
\begin{aligned}
\mathbf{X}_{i1} &\equiv \mathbf{y}_{i1} - \boldsymbol{\theta}_K \ln K_{i,0} - \boldsymbol{\theta}_1 = \mathbf{v}_i + \mathbf{e}_{i1} \\
&= \boldsymbol{\Gamma}_v \mathbf{z}_{i1} - \boldsymbol{\lambda}\widetilde{\boldsymbol{\rho}}'\widetilde{\mathbf{z}}_{i2} + \boldsymbol{\lambda}\varepsilon_{i1} + \widetilde{\mathbf{e}}_{i1} \\
\mathbf{X}_{it} &\equiv \Delta\mathbf{y}_{it} - \boldsymbol{\theta}_K \Delta \ln K_{i,t-1} - \Delta\boldsymbol{\theta}_t = \boldsymbol{\Gamma}_\eta \mathbf{z}_{it} + \mathbf{e}_{it} - \mathbf{e}_{i,t-1} \\
&= (\widetilde{\boldsymbol{\Gamma}}_\eta\widetilde{\boldsymbol{\rho}} - \boldsymbol{\lambda})\varepsilon_{i,t-1} + \boldsymbol{\lambda}\varepsilon_{it} + (\widetilde{\boldsymbol{\Gamma}}_\eta + \boldsymbol{\lambda}\widetilde{\boldsymbol{\rho}}')\widetilde{\mathbf{z}}_{it} - \boldsymbol{\lambda}\widetilde{\boldsymbol{\rho}}'\widetilde{\mathbf{z}}_{i,t+1} + \widetilde{\mathbf{e}}_{it} - \widetilde{\mathbf{e}}_{i,t-1}
\end{aligned}
$$

Now, define:

$$
\begin{aligned}
\mathbf{X}_i &= (\mathbf{X}_{i1}', ..., \mathbf{X}_{iT_i}')' \\
\boldsymbol{\varepsilon}_i &= (\varepsilon_{i1}', ..., \varepsilon_{iT_i}')' \\
\mathbf{z}_i &= (\mathbf{z}_{i1}', \widetilde{\mathbf{z}}_{i2}', ..., \widetilde{\mathbf{z}}_{i,T_i+1}')' \\
\mathbf{e}_i &= (\widetilde{\mathbf{e}}_{i1}', ..., \widetilde{\mathbf{e}}_{iT_i}')'.
\end{aligned}
$$

Then

$$
\begin{aligned}
\boldsymbol{\varepsilon}_i &\sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{dT_i}) \\
\mathbf{z}_i &\sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{r(T_i+1)}) \\
\mathbf{e}_i &\sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{T_i} \otimes \widetilde{\Sigma}_e)
\end{aligned}
$$

For firm i it is now possible to write the model as a simultaneous equation system as follows:

$$
\mathbf{X}_i = \mathbf{A}\mathbf{z}_i + \mathbf{B}\boldsymbol{\varepsilon}_i + \mathbf{C}\mathbf{e}_i
$$

where

$$
\mathbf{A} = \begin{bmatrix}
\boldsymbol{\Gamma}_v & -\boldsymbol{\lambda}\widetilde{\boldsymbol{\rho}}' & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & \widetilde{\boldsymbol{\Gamma}}_\eta + \boldsymbol{\lambda}\widetilde{\boldsymbol{\rho}}' & -\boldsymbol{\lambda}\widetilde{\boldsymbol{\rho}}' & \cdots & \mathbf{0} \\
\vdots & \cdots & \ddots & \ddots & \vdots \\
\mathbf{0} & \cdots & \mathbf{0} & \widetilde{\boldsymbol{\Gamma}}_\eta + \boldsymbol{\lambda}\widetilde{\boldsymbol{\rho}}' & -\boldsymbol{\lambda}\widetilde{\boldsymbol{\rho}}'
\end{bmatrix}
$$

$$
\mathbf{B} = \begin{bmatrix}
\boldsymbol{\lambda} & \mathbf{0} & \cdots & \mathbf{0} \\
\widetilde{\boldsymbol{\Gamma}}_\eta\widetilde{\boldsymbol{\rho}} - \boldsymbol{\lambda} & \boldsymbol{\lambda} & \cdots & \mathbf{0} \\
\vdots & \ddots & \ddots & \vdots \\
\mathbf{0} & \cdots & \widetilde{\boldsymbol{\Gamma}}_\eta\widetilde{\boldsymbol{\rho}} - \boldsymbol{\lambda} & \boldsymbol{\lambda}
\end{bmatrix}
$$

$$
\mathbf{C} = \begin{bmatrix} \mathbf{I}_d & \mathbf{0} & \cdots & \mathbf{0} \\ -\mathbf{I}_d & \mathbf{I}_d & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & -\mathbf{I}_d & \mathbf{I}_d \end{bmatrix}
$$

We see that $\mathbf{X}_i$ given z_i has the distribution

$$
\mathbf{X}_i | \mathbf{z}_i \sim \mathcal{N}(\mathbf{A}\mathbf{z}_i, \mathbf{B}\mathbf{B}' + \mathbf{C}(\mathbf{I}_{T_i} \otimes \widetilde{\Sigma}_e)\mathbf{C}')
$$

Although the dimension of the covariance matrix is dT_i, and hence increases with the number of observations T_i, there are no numerical problems involved here. The reason is that this matrix is block-tridiagonal, and hence its inversion requires only $O(T_i)$ operations (rather than the usual $O(T_i^3)$ for general matrices).

Since

$$
\left. \frac{\partial \ln f(\mathbf{X}_i, \boldsymbol{\varepsilon}_i | \mathbf{z}_i)}{\partial \boldsymbol{\varepsilon}_i} \right|_{\boldsymbol{\varepsilon}_i = E(\boldsymbol{\varepsilon}_i | \mathbf{X}_i, \mathbf{z}_i)} = \mathbf{0}
$$

$$
\mathrm{Var}(\boldsymbol{\varepsilon}_i | \mathbf{X}_i, \mathbf{z}_i) = - \left[\frac{\partial^2 \ln f(\mathbf{X}_i, \boldsymbol{\varepsilon}_i | \mathbf{z}_i)}{\partial \boldsymbol{\varepsilon}_i \partial \boldsymbol{\varepsilon}_i'} \right]^{-1} \equiv \boldsymbol{\Omega}
$$

we obtain

$$
\boldsymbol{\varepsilon}_i | (\mathbf{X}_i, \mathbf{z}_i) \sim \mathcal{N}(\boldsymbol{\Omega}\mathbf{B}' \left[\mathbf{C}(\mathbf{I}_{T_i} \otimes \widetilde{\Sigma}_e)\mathbf{C}' \right]^{-1} [\mathbf{X}_i - \mathbf{A}\mathbf{z}_i], \boldsymbol{\Omega})
$$

$$
\boldsymbol{\Omega}^{-1} = \mathbf{B}' \left[\mathbf{C}(\mathbf{I}_{T_i} \otimes \widetilde{\Sigma}_e)\mathbf{C}' \right]^{-1} \mathbf{B} + \mathbf{I}_{T_i}
$$

To obtain $\boldsymbol{\Omega}$, one needs to invert $\boldsymbol{\Omega}^{-1}$, which is non-sparse. However, from a well-known matrix inversion lemma, we have

$$
\boldsymbol{\Omega} = \left[\mathbf{B}' \left[\mathbf{C}(\mathbf{I}_{T_i} \otimes \widetilde{\Sigma}_e)\mathbf{C}' \right]^{-1} \mathbf{B} + \mathbf{I}_{T_i} \right]^{-1}
$$

$$
= \mathbf{I}_{T_i} - \mathbf{B}' \left[\mathbf{C}(\mathbf{I}_{T_i} \otimes \widetilde{\Sigma}_e)\mathbf{C}' + \mathbf{B}\mathbf{B}' \right]^{-1} \mathbf{B}
$$

which only requires inversion of the block-tridiagonal matrix $\mathbf{C}(\mathbf{I}_{T_i} \otimes \widetilde{\Sigma}_e)\mathbf{C}' + \mathbf{B}\mathbf{B}'$.

If we define

$$
\bar{\boldsymbol{\varepsilon}}_i = \left(\chi_{i1}(\boldsymbol{\gamma}_A' \mathbf{y}_{i1} + \boldsymbol{\gamma}_x \mathbf{x}_{i1} + \gamma_1), ..., \chi_{iT_i}(\boldsymbol{\gamma}_A' \mathbf{y}_{iT_i} + \boldsymbol{\gamma}_x \mathbf{x}_{iT_i} + \gamma_{T_i}) \right)'
$$

(note that $\chi_{iT_i} = 1$ for $t < T_i$) we obtain:

$$
f(\mathbf{X}_i, \chi_{iT_i} | \mathbf{z}_i) = f(\mathbf{X}_i | \mathbf{z}_i) \int_{\boldsymbol{\varepsilon}_i \leq \bar{\boldsymbol{\varepsilon}}_i} f(\boldsymbol{\varepsilon} | \mathbf{X}_i, \mathbf{z}_i) \, d\boldsymbol{\varepsilon}_i
$$

$$
\propto |\mathbf{B}\mathbf{B}' + \mathbf{C}(\mathbf{I}_{T_i} \otimes \widetilde{\Sigma}_e)\mathbf{C}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{X}_i - \mathbf{A}\mathbf{z}_i)' \left(\mathbf{B}\mathbf{B}' + \mathbf{C}(\mathbf{I}_{T_i} \otimes \widetilde{\Sigma}_e)\mathbf{C} \right)^{-1} (\mathbf{X}_i - \mathbf{A}\mathbf{z}_i) \right\} \times
$$

$$
\mathrm{Prob}(\boldsymbol{\nu}_i \leq \bar{\boldsymbol{\varepsilon}}_i - \boldsymbol{\Omega}\mathbf{B}' \left[\mathbf{C}(\mathbf{I}_{T_i} \otimes \widetilde{\Sigma}_e)\mathbf{C}' \right]^{-1} [\mathbf{X}_i - \mathbf{A}\mathbf{z}_i])
$$

$$
(33)
$$

where $\boldsymbol{\nu}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$. Computation of the multivariate normal probability in (33) is, of course, far from trivial. We apply the GHK smooth recursive simulator, which is both fast and simple to implement for our problem; see Hajivassiliou et al., 1996.

Finally, to obtain the likelihood function we need to "integrate out" $\mathbf{z}_i$:

$$
\begin{aligned}
f(\mathbf{X}_i, \chi_{iT_i}) &= \int f(\mathbf{X}_i, \chi_i | \mathbf{z}_i) \frac{1}{(\sqrt{2\pi})^{r(T_i+1)}} \exp(-\frac{1}{2} \mathbf{z}_i{}' \mathbf{z}_i) d\mathbf{z}_i \\
&\simeq \frac{1}{S} \sum_{s=1}^{S} f(\mathbf{X}_i, \chi_{iT_i} | \mathbf{z}_i^{(s)})
\end{aligned}
$$

where $z_i^{(s)}$ is the s'th random draw from the standard multivariate normal distribution for firm i, and S is the number of draws.

If we now define

$$
\begin{aligned}
l_i^{(s)}(\boldsymbol{\theta}) &= \ln \frac{f(\mathbf{X}_i, \chi_i | \mathbf{z}^{(s)})}{S} \\
l_i(\boldsymbol{\theta}) &= \ln \sum_{s=1}^{S} \exp(l_i^{(s)}(\boldsymbol{\theta})) \\
l(\boldsymbol{\theta}) &= \sum_i l_i(\boldsymbol{\theta})
\end{aligned}
$$

This gives

$$
\begin{aligned}
l(\boldsymbol{\theta}) &\simeq \sum_i \ln f(\mathbf{X}_i, \chi_i) \\
\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \sum_i \frac{\partial l_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_i \sum_s \frac{\exp(l_i^{(s)}(\boldsymbol{\theta}))}{\exp(l_i(\boldsymbol{\theta}))} \frac{\partial l_i^{(s)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\
&= \sum_i \sum_s \frac{\partial l_i^{(s)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \omega_i^{(s)}
\end{aligned}
$$

Although obtaining the derivatives $\frac{\partial l_i^{(s)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ are cumbersome, analytic expressions are available for all these derivatives by using well-know matrix-derivatives rules (see Lutkepohl (1996)). Hence an efficient quasi-Newton algorithm can be applied to solve the estimation problem.

Appendix B: Proof of equation (25)-(26).

From (14), (15) and (22), it follows that

$$
\boldsymbol{\Gamma}_\eta \mathbf{z}_{it} = \boldsymbol{\theta}_A \Delta \boldsymbol{\alpha}_{it}.
$$

Under the assumption of normality and independence of $\Delta \ln A_{it}^*$ and $(\Delta \ln c_{it}, \Delta \ln b_{it})$, we can write:

$$
\boldsymbol{\theta}_A \Delta \boldsymbol{\alpha}_{it} = \boldsymbol{\theta}_A \mathbf{S} \mathbf{D} \bar{\mathbf{z}}_{it}, \text{ where } \bar{\mathbf{z}}_{it} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_3),
$$

where $\mathbf{S}$ and $\mathbf{D}$ are defined in (23) and (26). Then, $\boldsymbol{\Gamma}_\eta = \boldsymbol{\theta}_A \mathbf{S} \mathbf{D} \mathbf{R}$, and $\mathbf{z}_{it} = \mathbf{R}' \bar{\mathbf{z}}_{it}$, for some orthonormal matrix $\mathbf{R}$ (see Anderson, 1984). We must show that $\mathbf{R} = \mathbf{I}_3$.

We have

$$\boldsymbol{\theta}_A \mathbf{S} = \begin{bmatrix} \frac{(e-1)}{\varepsilon+e-e\varepsilon}\sigma_A & -\frac{\varepsilon(e-1)}{\varepsilon+e-e\varepsilon}\sigma_c & 0 \\ \frac{(e-1)}{\varepsilon+e-e\varepsilon}\sigma_A & \left(-\frac{\varepsilon(e-1)}{\varepsilon+e-e\varepsilon}+\frac{\rho}{1-\rho}\right)\sigma_c & 0 \\ \frac{(e-1)}{\varepsilon+e-e\varepsilon}\sigma_A & \left(-\frac{\varepsilon(e-1)}{\varepsilon+e-e\varepsilon}+\frac{\rho}{1-\rho}\right)\sigma_c & \frac{\rho}{1-\rho}\sigma_b \\ (1-\phi)\kappa_A\sigma_A & (1-\phi)\kappa_c\sigma_c & 0 \end{bmatrix} \equiv \begin{bmatrix} \widetilde{\theta}_1 & \widetilde{\theta}_2 & 0 \\ \widetilde{\theta}_1 & \widetilde{\theta}_3 & 0 \\ \widetilde{\theta}_1 & \widetilde{\theta}_3 & \widetilde{\theta}_4 \\ \widetilde{\theta}_5 & \widetilde{\theta}_6 & 0 \end{bmatrix} \tag{34}$$

Let

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}.$$

Now consider the equation $\boldsymbol{\Gamma}_\eta = \boldsymbol{\theta}_A \mathbf{SDR}$:

$$\begin{bmatrix} \gamma_1 & \gamma_2 & 0 \\ \gamma_1 & \gamma_3 & 0 \\ \gamma_1 & \gamma_4 & \gamma_5 \\ \gamma_6 & \gamma_7 & 0 \end{bmatrix} = \begin{bmatrix} \widetilde{\theta}_1 & \widetilde{\theta}_2 & 0 \\ \widetilde{\theta}_1 & \widetilde{\theta}_3 & 0 \\ \widetilde{\theta}_1 & \widetilde{\theta}_3 & \widetilde{\theta}_4 \\ \widetilde{\theta}_5 & \widetilde{\theta}_6 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \sigma_{cb} & \sqrt{1-\sigma_{cb}^2} \end{bmatrix} \mathbf{R} \tag{35}$$

Since $\mathbf{R}$ is orthonormal, we easily find that $\mathbf{R} = \mathbf{I}_3$. Solving (35) with respect to $\widetilde{\theta}_1, ..., \widetilde{\theta}_6$ and σ_{cb}, then gives a unique solution when we require that $\sigma_{cb} < 0$ and $\widetilde{\theta}_2 < 0$ (which follows from (34)).

Appendix C: NACE sector codes

25 Manufacture of rubber and plastic products

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

30. Manufacture of office machinery and computers

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

34 Manufacture of motor vehicles, trailers, and semitrailers

35 Manufacture of other transport equipment

Appendix D: Some theoretical ideas on firm heterogeneity

We decompose the persistent differences in firm performance into (i) permanent differences that are established already when the firm enters an industry, and (ii) differences that are generated through subsequent, idiosyncratic innovations that accumulate through the firms' life-time[23]. In this appendix, we briefly review the main ideas in the theoretical literature emphasizing efficiency differences permanent to the firms and differences evolving through innovations that are cumulated, respectively.

The importance of permanent differences in efficiency: Which theoretical models can explain large permanent differences across firms that are introduced already when the firms enter the industry? An old idea is the so-called putty-clay model, emphasizing the irreversible nature of a firm's choice of technology. The classical contribution is Johansen (1959)[24]. The putty-clay literature emphasizes that choices of technology are embodied in the capital, which makes adjustment costly as it requires that the existing capital must be replaced.

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

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

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

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

---

[23]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 emanating 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).

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

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

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

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

Early research on firm heterogeneity was stimulated by Gibrat's analysis of the skewed size-distribution of firms, and how such skewed size-distributions can be generated from independent firm growth processes. These growth processes are characterized, according to the so-called Gibrat's law, by firm growth rates that are independent of firm size. Simon and his co-authors developed this line of research in the 1960s and 1970s, by exploring firm evolution through formal modelling of the stochastic processes; see Ijiri and Simon (1977). While this early work paid little attention to optimizing behavior and interactions between firms, Hopenhayn (1992) presents a related study of an industry equilibrium generated by interacting and optimizing firms. Firm growth is driven by exogenous stochastic processes, with exit as an endogenous decision[25].

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

---

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

Kortum (2004). These studies confront stylized facts that have emerged from a large number of empirical studies of R&D, innovation and firm growth.

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

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

Tables and figures

Table 1: **Descriptive statistics. Standard deviation in parenthesis**

Sector (NACE)	# of firms/entrants	Mean log of output*	Mean log of lab.prod**	# of empl.
Plastics (25)	367/126	7.78 (1.32)	-1.60 (.54)	13.8 (25.6)
Machinery (29)	1227/548	7.73 (1.34)	-1.61 (.52)	15.0 (32.2)
Electrical inst. (30-33)	785/342	7.76 (1.35)	-1.55 (.54)	16.5 (44.1)
Transp. eq. (34-35)	795/316	8.26 (1.45)	-1.65 (.46)	32.8 (94.5)

∗ Value added deflated by the consumer price index.
∗∗ Value added per hours worked.

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

Sector (NACE)	Eigenv. of Σ_η (Idiosyncratic innov.)	Eigenv. of Σ_v (Permanent differences)	Eigenv. of Σ_e (Noise)	Pseudo R^2
Plastics (25)	.233 (.021)	4.03 (.48)	.36	
	.033 (.002)	.22 (.04)	.09	
	.001 (.001)	.02 (.01)	.01	0.90
	.000 (.001)	.00 (.01)	.00	
Machinery (29)	.288 (.013)	4.10 (.29)	.50	
	.027 (.001)	.39 (.03)	.10	
	.002 (.001)	.03 (.01)	.02	0.88
	.000 (.001)	.00 (.01)	.01	
Electrical Inst. (30-33)	.220 (.019)	6.23 (.42)	.50	
	.023 (.003)	.42 (.04)	.06	
	.002 (.001)	.05 (.02)	.03	0.91
	.000 (.001)	.00 (.01)	.01	
Transp. Eq. (34-35)	.347 (.029)	4.82 (.33)	.45	
	.041 (.003)	.64 (.07)	.11	
	.005 (.001)	.04 (.01)	.04	0.91
	.000 (.001)	.01 (.01)	.01	

Table 3: **Estimates of factor loadings and capital coefficients in single factor model.** Standard errors in parenthesis.

Sector (NACE)	Idiosyn. inn.	Permanent dif.	Capital coef.	Pseudo R^2
	Estim. (st.dev.)	Estim. (st.dev.)	Estim. (st.dev.)	
Plastics (25)	.27(.02)	1.13 (.06)	.13 (.07)	0.89
	.26 (.05)	1.25 (.16)	.09 (.07)	
	.23 (.07)	1.02 (.18)	.15 (.08)	
	.11 (.03)	.37 (.08)	.72 (.05)	
Machinery (29)	.28(.02)	1.09 (.02)	.11 (.04)	0.86
	.29 (.03)	1.18 (.05)	.09 (.04)	
	.31 (.04)	1.18 (.06)	.10 (.04)	
	.10 (.02)	.33 (.01)	.70 (.03)	
Electrical Inst. (30-33)	.25(.01)	1.35 (.05)	.08 (.04)	0.89
	.25 (.03)	1.48 (.15)	.05 (.04)	
	.27 (.04)	1.40 (.20)	.05 (.04)	
	.08 (.02)	.50 (.02)	.64 (.03)	
Transp. Eq. (34-35)	.30(.03)	1.22 (.11)	.15(.06)	0.87
	.32(.11)	1.28 (.44)	.16 (.07)	
	.33 (.13)	1.26 (.50)	.17 (.08)	
	.08 (.02)	.36 (.06)	.71 (.03)	

Table 4: **Testing structural restrictions on loading coefficients and capital coefficients in single factor model.**

Sector (NACE)	Restrictions Γ_η			Restrictions Γ_v			Restrictions θ_K		
	χ^2	d.f.	P-value	χ^2	d.f.	P-value	χ^2	d.f.	P-value
Plastics (25)	.46	2	.79	.26	2	.87	2.74	2	.25
Machinery (29)	.03	2	.98	2.51	2	.28	9.00	2	.01
Electrical Inst. (30-33)	.20	2	.90	7.69	2	.02	4.21	2	.12
Transp. Eq. (34-35)	.48	2	.78	2.02	2	.36	.25	2	.87

Table 5: **Estimates of the extended model and test of overidentifying restrictions.**

Sector (NACE)	Parameter estimates							χ^2-test
	$\widetilde{\theta}_1$	$\widetilde{\theta}_2$	$\widetilde{\theta}_3$	$\widetilde{\theta}_4$	$\widetilde{\theta}_5$	$\widetilde{\theta}_6$	σ_{cb}	P-value
Plastics (25)	.26	−.07	.02	.22	.10	−.04	−.63	.12
Machinery (29)	.28	−.01	−.09	−.19	.10	.00	−.57	.08
Elect. In. (30-33)	.25	−.01	−.08	−.21	.07	.00	−.58	.09
Tran. Eq. (34-35)	.30	−.05	−.15	−.26	.07	−.01	−.57	.17

Table 6: **Measures of the origins of firm heterogeneity.** The unconditional variance ratio (V), the conditional variance ratio (CV) and the relative variance (RV). Estimates from the extended model.

Sector (NACE)	$T^* = \frac{\text{tr } Var(\Sigma_{\mathbf{v}})}{\text{tr } Var(\Sigma_\eta)}$	V	CV	RV
Plastics (25)	19.2	2.04	1.68	.79
Machinery (29)	12.5	1.70	1.62	.82
Electr. inst. (30-33)	31.1	3.45	3.04	.77
Transp. eq. (34-35)	16.1	1.76	1.58	.70

Table 7: **Testing of the missing at random assumption**

Sector (NACE)	Zero restrictions ρ			Zero restrictions λ		
	χ^2	d.f.	P-value	χ^2	d.f.	P-value
Plastics (25)	5.74	1	.02	1.16	4	.88
Machinery (29)	12.0	1	.00	2.68	4	.61
Electrical Inst. (30-33)	7.81	1	.00	4.96	4	.30
Transp. Eq. (34-35)	3.70	1	.05	8.25	4	.09
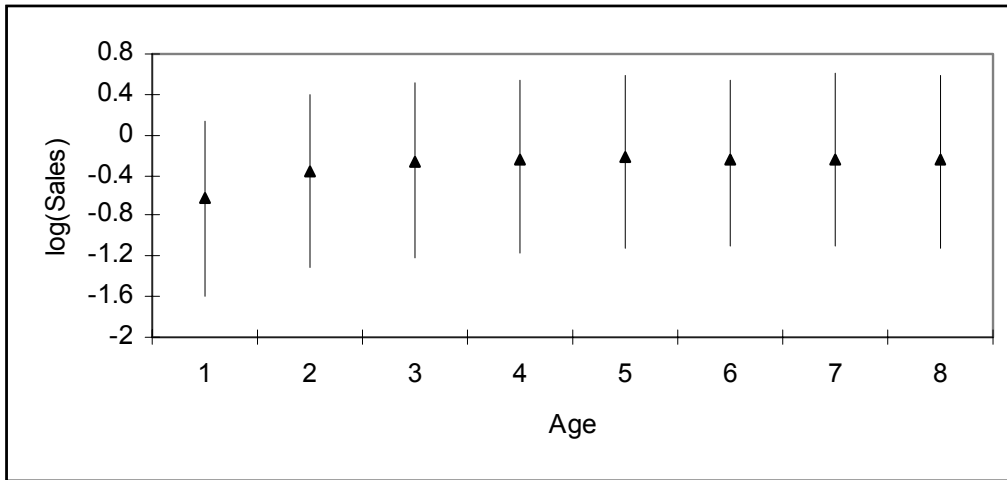
Figure 1: **Differences in log sales as a function of firm age.** Triangles indicates the means and the vertical lines show the interquartile range.
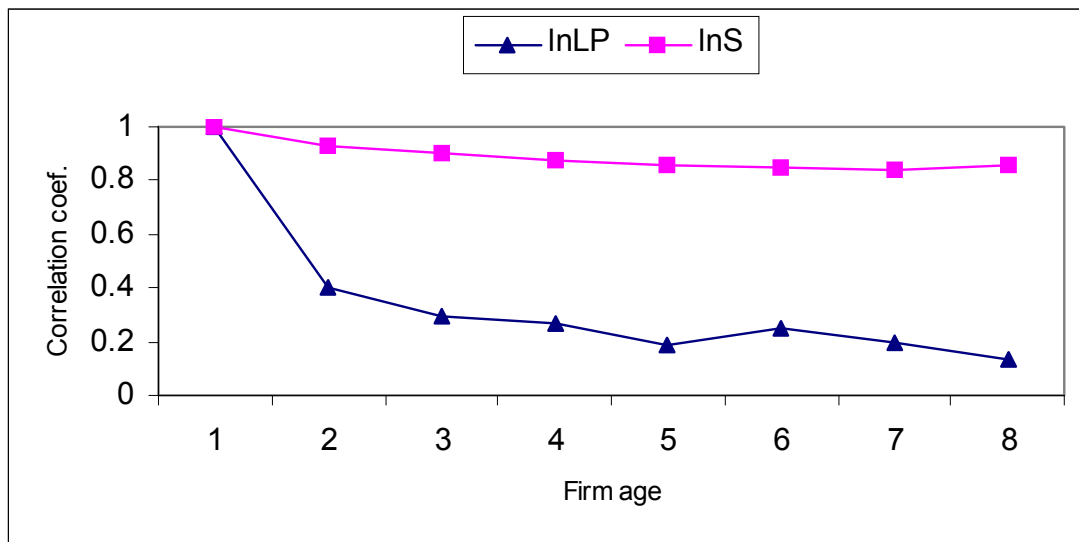

Figure 2: **The correlation between relative performance in a firm's first year and in its subsequent years.** The squares correspond to the correlation coefficents for (log) sales while the triangles refer to (log) labor productivity.
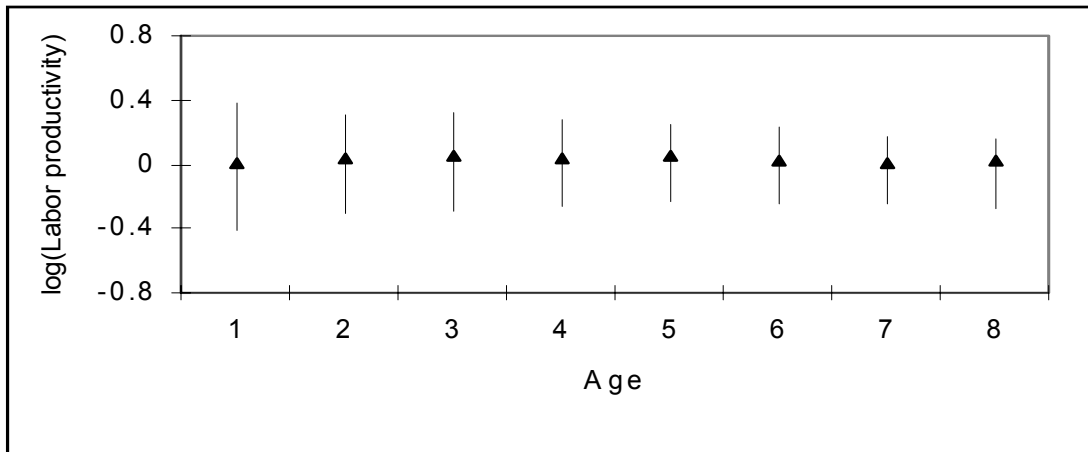
Figure 3: **Differences in log labor productivity as a function of firm age.** Triangles indicate the means and the vertical lines show the interquartile range.
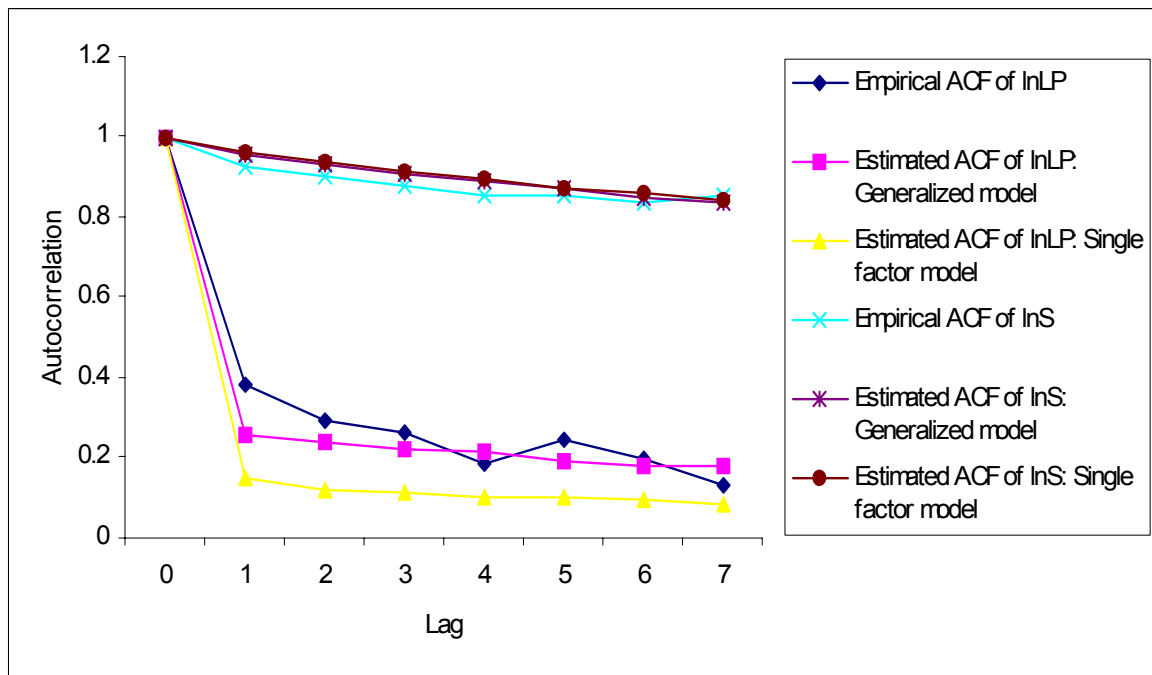

Figure 4: **Autocorrelation functions (ACF) for log of labor productivity (lnLP) and log of sales (lnS).**

Recent publications in the series Discussion Papers

310   M. Rege and K. Telle (2001): An Experimental Investigation of Social Norms

311   L.C. Zhang (2001): A method of weighting adjustment for survey data subject to nonignorable nonresponse

312   K. R. Wangen and E. Biørn (2001): Prevalence and substitution effects in tobacco consumption. A discrete choice analysis of panel data

313   G.H. Bjertnær (2001): Optimal Combinations of Income Tax and Subsidies for Education

314   K. E. Rosendahl (2002): Cost-effective environmental policy: Implications of induced technological change

315   T. Kornstad and T.O. Thoresen (2002): A Discrete Choice Model for Labor Supply and Child Care

316   A. Bruvoll and K. Nyborg (2002): On the value of households' recycling efforts

317   E. Biørn and T. Skjerpen (2002): Aggregation and Aggregation Biases in Production Functions: A Panel Data Analysis of Translog Models

318   Ø. Døhl (2002): Energy Flexibility and Technological Progress with Multioutput Production. Application on Norwegian Pulp and Paper Industries

319   R. Aaberge (2002): Characterization and Measurement of Duration Dependence in Hazard Rate Models

320   T. J. Klette and A. Raknerud (2002): How and why do Firms differ?

321   J. Aasness and E. Røed Larsen (2002): Distributional and Environmental Effects of Taxes on Transportation

322   E. Røed Larsen (2002): The Political Economy of Global Warming: From Data to Decisions

323   E. Røed Larsen (2002): Searching for Basic Consumption Patterns: Is the Engel Elasticity of Housing Unity?

324   E. Røed Larsen (2002): Estimating Latent Total Consumption in a Household.

325   E. Røed Larsen (2002): Consumption Inequality in Norway in the 80s and 90s.

326   H.C. Bjørnland and H. Hungnes (2002): Fundamental determinants of the long run real exchange rate:The case of Norway.

327   M. Søberg (2002): A laboratory stress-test of bid, double and offer auctions.

328   M. Søberg (2002): Voting rules and endogenous trading institutions: An experimental study.

329   M. Søberg (2002): The Duhem-Quine thesis and experimental economics: A reinterpretation.

330   A. Raknerud (2002): Identification, Estimation and Testing in Panel Data Models with Attrition: The Role of the Missing at Random Assumption

331   M.W. Arneberg, J.K. Dagsvik and Z. Jia (2002): Labor Market Modeling Recognizing Latent Job Attributes and Opportunity Constraints. An Empirical Analysis of Labor Market Behavior of Eritrean Women

332   M. Greaker (2002): Eco-labels, Production Related Externalities and Trade

333   J. T. Lind (2002): Small continuous surveys and the Kalman filter

334   B. Halvorsen and T. Willumsen (2002): Willingness to Pay for Dental Fear Treatment. Is Supplying Fear Treatment Social Beneficial?

335   T. O. Thoresen (2002): Reduced Tax Progressivity in Norway in the Nineties. The Effect from Tax Changes

336   M. Søberg (2002): Price formation in monopolistic markets with endogenous diffusion of trading information: An experimental approach

337   A. Bruvoll og B.M. Larsen (2002): Greenhouse gas emissions in Norway. Do carbon taxes work?

338   B. Halvorsen and R. Nesbakken (2002): A conflict of interests in electricity taxation? A micro econometric analysis of household behaviour

339   R. Aaberge and A. Langørgen (2003): Measuring the Benefits from Public Services: The Effects of Local Government Spending on the Distribution of Income in Norway

340   H. C. Bjørnland and H. Hungnes (2003): The importance of interest rates for forecasting the exchange rate

341   A. Bruvoll, T.Fæhn and Birger Strøm (2003): Quantifying Central Hypotheses on Environmental Kuznets Curves for a Rich Economy: A Computable General Equilibrium Study

342   E. Biørn, T. Skjerpen and K.R. Wangen (2003): Parametric Aggregation of Random Coefficient Cobb-Douglas Production Functions: Evidence from Manufacturing Industries

343   B. Bye, B. Strøm and T. Åvitsland (2003): Welfare effects of VAT reforms: A general equilibrium analysis

344   J.K. Dagsvik and S. Strøm (2003): Analyzing Labor Supply Behavior with Latent Job Opportunity Sets and Institutional Choice Constraints

345   A. Raknerud, T. Skjerpen and A. Rygh Swensen (2003): A linear demand system within a Seemingly Unrelated Time Series Equation framework

346   B.M. Larsen and R.Nesbakken (2003): How to quantify household electricity end-use consumption

347   B. Halvorsen, B. M. Larsen and R. Nesbakken (2003): Possibility for hedging from price increases in residential energy demand

348   S. Johansen and A. R. Swensen (2003): More on Testing Exact Rational Expectations in Cointegrated Vector Autoregressive Models: Restricted Drift Terms

349   B. Holtsmark (2003): The Kyoto Protocol without USA and Australia - with the Russian Federation as a strategic permit seller

350   J. Larsson (2003): Testing the Multiproduct Hypothesis on Norwegian Aluminium Industry Plants

351   T. Bye (2003): On the Price and Volume Effects from Green Certificates in the Energy Market

352   E. Holmøy (2003): Aggregate Industry Behaviour in a Monopolistic Competition Model with Heterogeneous Firms

353   A. O. Ervik, E.Holmøy and T. Hægeland (2003): A Theory-Based Measure of the Output of the Education Sector

354   E. Halvorsen (2003): A Cohort Analysis of Household Saving in Norway

355   I. Aslaksen and T. Synnestvedt (2003): Corporate environmental protection under uncertainty

356   S. Glomsrød and W. Taoyuan (2003): Coal cleaning: A viable strategy for reduced carbon emissions and improved environment in China?

357 A. Bruvoll T. Bye, J. Larsson og K. Telle (2003): Technological changes in the pulp and paper industry and the role of uniform versus selective environmental policy.

358 J.K. Dagsvik, S. Strøm and Z. Jia (2003): A Stochastic Model for the Utility of Income.

359 M. Rege and K. Telle (2003): Indirect Social Sanctions from Monetarily Unaffected Strangers in a Public Good Game.

360 R. Aaberge (2003): Mean-Spread-Preserving Transformation.

361 E. Halvorsen (2003): Financial Deregulation and Household Saving. The Norwegian Experience Revisited

362 E. Røed Larsen (2003): Are Rich Countries Immune to the Resource Curse? Evidence from Norway's Management of Its Oil Riches

363 E. Røed Larsen and Dag Einar Sommervoll (2003): Rising Inequality of Housing? Evidence from Segmented Housing Price Indices

364 R. Bjørnstad and T. Skjerpen (2003): Technology, Trade and Inequality

365 A. Raknerud, D. Rønningen and T. Skjerpen (2003): A method for improved capital measurement by combining accounts and firm investment data

366 B.J. Holtsmark and K.H. Alfsen (2004): PPP-correction of the IPCC emission scenarios - does it matter?

367 R. Aaberge, U. Colombino, E. Holmøy, B. Strøm and T. Wennemo (2004): Population ageing and fiscal sustainability: An integrated micro-macro analysis of required tax changes

368 E. Røed Larsen (2004): Does the CPI Mirror Costs.of.Living? Engel's Law Suggests Not in Norway

369 T. Skjerpen (2004): The dynamic factor model revisited: the identification problem remains

370 J.K. Dagsvik and A.L. Mathiassen (2004): Agricultural Production with Uncertain Water Supply

371 M. Greaker (2004): Industrial Competitiveness and Diffusion of New Pollution Abatement Technology – a new look at the Porter-hypothesis

372 G. Børnes Ringlund, K.E. Rosendahl and T. Skjerpen (2004): Does oilrig activity react to oil price changes? An empirical investigation

373 G. Liu (2004) Estimating Energy Demand Elasticities for OECD Countries. A Dynamic Panel Data Approach

374 K. Telle and J. Larsson (2004): Do environmental regulations hamper productivity growth? How accounting for improvements of firms' environmental performance can change the conclusion

375 K.R. Wangen (2004): Some Fundamental Problems in Becker, Grossman and Murphy's Implementation of Rational Addiction Theory

376 B.J. Holtsmark and K.H. Alfsen (2004): Implementation of the Kyoto Protocol without Russian participation

377 E. Røed Larsen (2004): Escaping the Resource Curse and the Dutch Disease? When and Why Norway Caught up with and Forged ahead of Its Neughbors

378 L. Andreassen (2004): Mortality, fertility and old age care in a two-sex growth model

379 E. Lund Sagen and F. R. Aune (2004): The Future European Natural Gas Market - are lower gas prices attainable?

380 A. Langørgen and D. Rønningen (2004): Local government preferences, individual needs, and the allocation of social assistance

381 K. Telle (2004): Effects of inspections on plants' regulatory and environmental performance - evidence from Norwegian manufacturing industries

382 T. A. Galloway (2004): To What Extent Is a Transition into Employment Associated with an Exit from Poverty

383 J. F. Bjørnstad and E.Ytterstad (2004): Two-Stage Sampling from a Prediction Point of View

384 A. Bruvoll and T. Fæhn (2004): Transboundary environmental policy effects: Markets and emission leakages

385 P.V. Hansen and L. Lindholt (2004): The market power of OPEC 1973-2001

386 N. Keilman and D. Q. Pham (2004): Empirical errors and predicted errors in fertility, mortality and migration forecasts in the European Economic Area

387 G. H. Bjertnæs and T. Fæhn (2004): Energy Taxation in a Small, Open Economy: Efficiency Gains under Political Restraints

388 J.K. Dagsvik and S. Strøm (2004): Sectoral Labor Supply, Choice Restrictions and Functional Form

389 B. Halvorsen (2004): Effects of norms, warm-glow and time use on household recycling

390 I. Aslaksen and T. Synnestvedt (2004): Are the Dixit-Pindyck and the Arrow-Fisher-Henry-Hanemann Option Values Equivalent?

391 G. H. Bjønnes, D. Rime and H. O.Aa. Solheim (2004): Liquidity provision in the overnight foreign exchange market

392 T. Åvitsland and J. Aasness (2004): Combining CGE and microsimulation models: Effects on equality of VAT reforms

393 M. Greaker and Eirik. Sagen (2004): Explaining experience curves for LNG liquefaction costs: Competition matter more than learning

394 K. Telle, I. Aslaksen and T. Synnestvedt (2004): "It pays to be green" - a premature conclusion?

395 T. Harding, H. O. Aa. Solheim and A. Benedictow (2004). House ownership and taxes

396 E. Holmøy and B. Strøm (2004): The Social Cost of Government Spending in an Economy with Large Tax Distortions: A CGE Decomposition for Norway

397 T. Hægeland, O. Raaum and K.G. Salvanes (2004): Pupil achievement, school resources and family background

398 I. Aslaksen, B. Natvig and I. Nordal (2004): Environmental risk and the precautionary principle: "Late lessons from early warnings" applied to genetically modified plants

399 J. Møen (2004): When subsidized R&D-firms fail, do they still stimulate growth? Tracing knowledge by following employees across firms

400 B. Halvorsen and Runa Nesbakken (2004): Accounting for differences in choice opportunities in analyses of energy expenditure data

401 T.J. Klette and A. Raknerud (2004): Heterogeneity, productivity and selection: An empirical study of Norwegian manufacturing firms