Forecasting productivity using information from firm-level data

Zoltán Wolf[1]

Supervised (and powered) by: Prof. Eric J. Bartelsman

August 2008

[1]I am grateful to Prof. E.J. Bartelsman for all his help and support. Part of this research has been carried out at the Centrum voor Beleidstatistiek, of which the help I would like to acknowledge. I would also like to thank Dave Reifschneider of the Washington Fed, Jerome Henry of the ECB and Joris de Wind of the DNB for providing information on current central bank practices.

Abstract

This paper contributes to this vast literature by tying together work from the firm-level productivity studies with macro-level efforts to forecast productivity growth. While the paper spends considerable effort to measure productivity at the firm-level, the main question is *how* to use information available at the firm level to improve macro-level productivity forecasts. To our knowledge, our study is a novel attempt to connect micro and macro level analysis whereby micro-level productivity estimates and decompositions of aggregate productivity provide additional information to be used in making macro forecasts.

We carried out several forecasting experiment for different specifications and models. Although there are some remaining questions, results suggest that micro-aggregated decompositions improve aggregate total factor productivity forecasts.

JEL classification: C14, C23, D24, O12, O47

Keywords: economic growth, production function, total factor productivity, aggregation, panel data

Contents

List of Tables

List of Figures

1 Introduction

Productivity and technical change have long been at the core of economic research. However, they received special attention in the last decade owing to the productivity boom in the second half of the nineties, and led to a large number of studies, research meetings and conferences on modeling, measuring, and interpreting productivity at various levels of aggregation.

This paper contributes to this vast literature by tying together work from the firm-level productivity studies with macro-level efforts to forecast productivity growth. While the paper spends considerable effort to measure productivity at the firm-level, the main question is *how* to use information available at the firm level to improve macro-level productivity forecasts. To our knowledge, our study is a novel attempt to connect micro and macro level analysis whereby micro-level productivity estimates and decompositions of aggregate productivity provide additional information to be used in making macro forecasts.

Although efforts to measure total factor productivity both at the micro and the macro level have posted remarkable results regarding several important issues - to alleviate measurement error problems, accounting for adjustment costs, variable factor utilization, intangibles, etc (e.g. Oliner et al.2007, Kátay and Wolf2008); prices and issues connected to aggregation (e.g. Petrin and Levinsohn2008) -, much less progress has been made in the practical matter of forecasting aggregate productivity.

A large number of practitioners face the generally formidable task of having to provide forecasts. For well-understood variables and systems this is less demanding in general. However, estimation and forecasting is quite complicated for tfp. By definition, tfp is the ratio between output and inputs. Measurement errors in either the numerator or the denominator do not cancel but exacerbate each other. A common finding of the productivity literature is that the yearly growth rates of tfp are often unstable or erratic, which makes extrapolation difficult.[1]

Further, theory is not unambiguous to the level of analysis at which tfp movement takes place. Are explicit actions at the firm level the source of movements in tfp and does aggregate tfp move owing to within- and between-firm movements in productivity and market share, or are representative firms merely able to pluck the fruits of movements in some exogenous productivity frontier (as is assumed in the RBC literature)? Evidence clearly points to the former (Bartelsman and Doms2000, Bartelsman et al.2005). The issue of understanding aggregate productivity movements thus requires good firm-level measures and proper theory on aggregation. Two seminal contributions on aggregation were Domar1961 and Hulten1978, while measurement of firm-level tfp has been explored by Levinsohn and Petrin2003, earlier Olley and Pakes1996 and a recent study by Petrin and Levinsohn2008.

The current practice of forecasting tfp at the macro level continues to be based on the representative firm view and thus only needs to worry about measurement at the aggregate level and does not need to consider firm-level heterogeneity, technology diffusion, shifts in market share, and appropriate aggregation methods. Tfp enters macroeconometric models typically through a production function equation, which is used to estimate potential output.

[1] For this reason, most studies report annual percent changes calculated over several years depending on the dataset (for example: Corrado et al.2006 and Oliner et al.2007).

However, concrete applications differ as to how tfp is modeled. For instance, the Federal Reserve's model for the US economy (FRB/US) models it as a stochastic process with shocks to both its level and growth rate. Trend tfp is estimated conditional on the values of other trends that enter the calculation of potential output[2] and eventually potential hours worked. The estimation tool is the Kalman filter as tfp is unobserved. Its trend is then extrapolated into the future assuming no future shocks to the level or growth rate.

The practice is different at the ECB (see Fagan et al.2001). The same framework (Cobb-Douglas production function with calibrated factor shares) is used to derive theoretically consistent first-order conditions that enter other equations in the model. In-sample trend tfp-values are calculated by applying the Hodrick-Prescott filter to the residual of the production function. However, the ECB does not directly forecast tfp.[3]

The objective in this paper is to improve on extrapolations of macro-level tfp. The forecasting idea is fairly simple: instead of looking only at the aggregate timeseries, we exploit movements over time in the underlying micro-level distributions of productivity and market share. These movements are captured by the timeseries of components of so-called productivity decompositions (Olley and Pakes1996, Baily et al.2001). These decompositions consider aggregate productivity essentially as an average of firm-level productivity measures weighted by appropriate shares. We also introduce and estimate a simple auxiliary equation for both individual productivity changes and individual shares. The objective of the auxiliary models is to extract information in the evolving joint cross-sectional distribution of firm-level productivity and firm-level share that can help in forecasting aggregate productivity.

We expect these components of productivity decompositions to help better gauge aggregate productivity developments. We believe that cutting the aggregate along certain dimensions may result in components that show sta-

[2]via a Cobb-Douglas froduction function

[3]To our present knowledge, the Dutch National Bank systematically forecasts GDP growth and inflation but not productivity separately.

bility and hence forecasting power. Our dataset, provided by Centrum voor Beleidsstatistiek[4], consists of a large panel of Dutch manufacturing firms. More details can be found in the data appendix.

2 Firm-level tfp measures

We applied a standard procedure developed by Levinsohn and Petrin2003 (LP henceforth). We present the estimation method only briefly, for further discussion and remaining issues in these techniques, see Ackerberg and Caves2004.

2.1 Outline of the method

Consider a Cobb-Douglas production function (indices i and t were dropped for simplification)

$$y = \beta_0 + \beta_l l + \beta_k k + \tau + \varepsilon \tag{1}$$

where y is log value-added, k is log capital, l is log labour, τ is a productivity shock and ε is assumed to be an *iid* disturbance. τ is unobservable by the econometrician but known[5] to the decision-maker. Since τ is in the information set of the firm on which it conditions its optimal choices of inputs, there will always be a nonnegative correlation between input factors and τ. This dependence renders simple OLS parameter estimates to be biased.

One of the key assumptions in LP is that capital is fixed, ie its level is chosen before production takes place. Hence, the orthogonality of k to the innovation in τ can be used to identify β_k. To solve the endogeneity problem with respect to freely variable labor, both estimation methods make use of a proxy. The key assumption is that the proxy is monotonic in τ. This is indispensable because the proxy is used to invert out the unobserved productivity shock.

[4] Dutch Statistics Office

[5] at least up to its expected value

There have been previous instances of this idea in the literature. Olley and Pakes1996 assumed that optimal investment is a strictly increasing function of current productivity: $i = i(\tau, k)$. If investment is strictly monotonous in τ then tfp can be solved for: $\tau = i^{-1}(\tau, k)$, where the inversion takes the form of a polynomial in i and k. Then the polynomial approximation of τ is used as the productivity proxy and, eventually this non-parametric function is used to substitute for tfp in the production function. LP criticized the investment proxy on several grounds. First, it is known that capital adjustment is lumpy because of nonconvexities. The sluggish response of capital to observed productivity shocks can easily violate the strict monotonity condition. Second, lumpiness also implies that zero investment observations have to be truncated from the data to preserve strict monotonicity. This leads to efficiency loss and to endogenous selection bias that one should control for. Therefore, LP suggest to use intermeriate inputs as a proxy: $m = m(\tau, k) \Longrightarrow \tau = m^{-1}(m, k)$.

One can argue for the validity of this proxy as follows. Higher productivity implies higher marginal product for capital for price-taking firms. As a response, firms increase production until the marginal product of capital equals its rental rate. To increase output, firms increase all inputs including intermediate inputs. Therefore, a change in intermediate input usage informs us about the change in productivity. LP's second criticism about the possible negativity of investment is absent here as intermediate input consumption is hardly ever negative.

Technically, the first step is to run a simple OLS regression of value-added on labor and the nonparametric function $m^{-1}(m, k)$.[6] In the estimation

[6]In principle, we should include output prices relative to input material/investment prices $(p^y - p^m)$ in the input demand fuction because firms' input consumption or investment might increase without positive productivity shocks just because output prices might rise faster than prices of input materials or investment. Therefore, it might be optimal to adjust the capital stock or increase input consumption even in absence of a productivity shock. Unfortunately, we did not have data for these prices but a better specification should incorporate price effects as well. Note that not including prices amounts to assuming input material and output prices are constant across firms.

process, we used a third order polinomial approximation of m^{-1} with fixed parameters across firms and time. In the above regression $\hat{\beta}_l$ is identified but $\hat{\beta}_k$ is not. This is because the polynomial $\phi(m,k)$ contains capital terms so k is collinear with the non-parametric function:

$$y = \beta_l l + \phi(m,k) + \varepsilon, \tag{2}$$

where $\phi(m,k) = \beta_0 + m^{-1}(m,k) + \beta_k k$. Hence, one obtains an estimate of $\hat{\beta}_l$ and $\hat{\phi}$. The problem is that $\hat{\phi}$ is not separable in $\hat{\beta}_k$.

The second step identifies capital's parameter, β_k. Here LP assume that τ follows a first order Markov process: $\tau_t = E[\tau_t|\tau_{t-1}] + \xi_t$, where the conditional expectation $E[\tau_t|\tau_{t-1}]$ is obtained from a nonparametric regression of $\left(\hat{\phi}_t - \beta_k k\right)$ on its lagged values. In short, an initial value of $\beta_k k$ permits to obtain τ_t, $E[\tau_t|\,\tau_{t-1}]$ and thus $\xi_t(\beta_k)$. As capital is fixed, the orthogonality of innovation ξ_t and k_t provides a moment condition: $E[\xi_t|\,k_t] = 0$. Additional moment conditions are also available as innovations should also be uncorrelated with lagged values of labour or itermediate materials.[7] To estimate standard errors a bootstrap procedure is used.

2.2 Production function parameter estimates

Production function parameter estimates are presented in tables 3 and 4. Column 1 shows the industry code - and table 1 gives the legend to them -, columns 2-4 contain the point estimates for K, L and M, column 5 lists

[7] In Olley-Pakes (1996), a bit stricter assumption about the data generating process of productivity allows a simpler algorithm. Assume that TFP follows a random walk. Then, the linear projection of output in excess of freely variable inputs can be written as

$$E\left[y_t - \hat{\beta}_l l_t\right] = \beta_0 + \beta_k k_t + E[\tau_t|\,\tau_{t-1}] = \beta_0 + \beta_k k_t + \hat{\phi}_{t-1} - \beta_k k_{t-1} - \beta_0 \tag{3}$$

$$E\left[y_t - \hat{\beta}_l l_t\right] = \beta_k (k_t - k_{t-1}) + \hat{\phi}_{t-1} \tag{4}$$

That is, if use the random walk assumption, all we need to do to get $\hat{\beta}_k$ is to regress $\left(y_t - \hat{\beta}_l l_t\right)$ on $(k_t - k_{t-1})$ and $\hat{\phi}_{t-1}$ without a constant.

the value of a χ^2 test statistic for constant returns to scale, columns 5-6 show standard errors for $\widehat{\beta}_k$ and $\widehat{\beta}_l$, and the last column shows the number of firm-year observations. The output and value added models seem to be in line in terms of ranking of industries according to production function parameter estimates. The $\widehat{\beta}_k$ are small in the output model relative to the value added model, which was against our expectations. Since the output specification requires a more complex IV technique together with some grid-search algorithm, it is less clear what can explain the difference between the two models. Therefore, in the sequel, we focus mostly on value-added-based tfp.

2.3 Caveats of Levinsohn-Petrin's method

Even though considered to be the core of state-of-the-art production function estimation, the methodology leaves several open questions. First, it is not straightforward that intermediate input material consumption should increase in response to a positive productivity shock at all times. One can think of cases when a positive productivity shock is associated with constant or even decreasing input usage. For instance, firms can improve their productivity by reducing intermediate input consumption by improving the quality control system and thus reducing the number of defective items produced (see Javocik2004).

Some of the identification assumptions of the procedures need further attention. It is crucial to assume that firms decide on labour having decided about the proxy. Otherwise – first labour then proxy – labour would be part of the state space on which the optimal amount of proxy is determined, ie. l would enter the input materials/investment demand functions. Technically, this would imply that the m function would contain l as an argument and that the inverse function m^{-1} would also contain l terms. This, in turn, would mean labor would be collinear with the polynomial approximation

$m^{-1}(m,k,l)$ and as an obvious consequence, the labor parameter could not be identified separately from this polynomial.[8]

It turns out that even the strict timing requirement is not enough to render the identification realistic. One also needs to assume that productivity changes between the decisions about the proxy and labour. If it does not, then it is as if the two were determined on the same information set. However, if tfp evolves between the two decisions then labour responds but the shocks are not controlled for by the proxy since that has been already decided upon. In other words, the uncontrolled productivity innovation feeds into the error term and since labour responds by assumption, labour and the cross product are correlated with the error term rendering their first-stage parameter estimates upwardly biased.

Despite all these criticisms, we think the method is worth the time. It is structured, intuitive and relatively well tractable. Moreover, as laid out in the introduction, our forecasting exercises do not concern measurement issues in the first place.

3 Forecasting

3.1 Methodology

As mentioned in the introduction, yearly growth rates of total factor productivity are often volatile. A common way to cope with huge variation in measured yearly growth rates is to analyze changes over a longer period. As an illustration, figure 1 shows yearly tfp-changes and table 5 shows average percent per annum changes for five-year periods for several countries.

Figure 2 depicts our micro-based and 'published' tfp measures for aggregate Dutch manufacturing. Micro-based aggregates are built using firm-level productivity estimates, 'published' aggregates come from the EUKLEMS

[8] This is the same story LP and OP tells about the capital parameter and this is the reason why the capital parameter can only be obtained through a two-step procedure.

database (O'Mahony et al.2007). We will refer to our firm-based tfp aggregates as micro-aggregated tfp hereafter. While the published EUKLEMS aggregates are built from national accounts data, which in turn capture information from the same firm-level data used in this paper, the published and micro-aggregated tfp measures differ for a variety of reasons relating to statistical practice. To trace these differences to underlying measures of inputs and outputs, figures 3, 4 and 5 graph year-on-year growth rates for value added, output and labor. Several stylized facts emerge from these graphs.

First, aggregate yearly growth rates exhibit large variation in all countries (figure 1). Second, micro-aggregated measures generally vary more than published EUKLEMS-based aggregates (figures 3, 4, 5). Third, micro-aggregated dynamics are similar on the input and output side of the firm suggesting that yearly tfp growth numbers might actually make sense despite their variation.[9] Finally, micro-aggregated tfp-growth series - although more volatile than that from EUKLEMS - capture the general dynamic pattern of the EUKLEMS aggregate reasonably well (figure 2).

The most prominent of all the above observations is that yearly aggregate productivity growth exhibits large variation. Therefore, following the best practice, we evaluate our estimates in 3 or 5-year averages instead of focusing on yearly forecasts. It also implies that we are not going to calculate standard MSE-type values to compare different forecasts but simply compare growth rates implied by fitted or extrapolated values.

Having seen some stylized facts, we can now turn the forecasting methodology. For practitioners, an obvious choice to forecast an aggregate timeseries is to fit a simple model using standard timeseries techniques. Our approach is different: we exploit standard techniques to estimate simple timeseries models for our micro-aggregated decompositions. In so doing, we investigate whether we can improve on the performance of these models both per se and relative to the forecast of the 'published' aggregate. Hereafter, we focus on the overall manufacturing aggregate if not stated otherwise.

[9] The same conclusion emerges from industry-by-industry analysis, not shown here. Micro-macro discrepancies are all the more obvious looking at level charts.

We make use of decompositions developed in 3.2, to reveal components that facilitate interpreting the dynamics of aggregate productivity. They consist of three building blocks, each showing the contribution to aggregate productivity growth. First, the within component shows the effects of firm-level productivity change. Second, the between component shows how the reallocation of resources among firms contributes to aggregate productivity growth. The third component encapsulates the effects of firm entry and exit.

We also estimate and introduce two auxiliary models for firm-level productivity growth and firm-level shares. In so doing we aim at gaining firm-level productivity and share measures with higher signal-to-noise ratio by using the fitted values from the auxiliary models.

We organize our analysis along four dimensions. The first dimension gives two alternatives because our firm-level productivity measures can be based on either output or value added (details in Levinsohn-Petrin2003). Second, since shares can be calculated using either the input or output-side variables of the firm, we have two weight options to build aggregates. We can use value added or output on the output-side and a weighted average of primary inputs on the input side. The weighted average of inputs is the fitted value of the production function. The third dimension also gives two alternatives because we evaluate two decompositions. Finally, we have the option of using ($\Delta\tau_{it}$ and ϕ_{it}) or ($\widehat{\Delta\tau}_{it}$ and $\widehat{\phi}_{it}$) because our auxiliary models yield fitted values for these. These alternatives leave us with 16 different scenarios.

In the forecasting exercises below, our target is always the average trend productivity growth of the published aggregate over the forecast horizon. We measure trend productivity by the Hodrick-Prescott trend[10] of the published EUKLEMS aggregate. The first experiment takes the forecast of the aggregate to be the extrapolated trend growth rate from a regression of EUKLEMS-aggregate productivity on a time-trend and a constant. The forecast of a component (of the two decompositions) is given by the dynamic

[10] which is observed as we ran regressions on a smaller sample

extrapolation from a simple autoregressive model[11]. The micro-aggregated forecast is the sum of the component-forecasts. Forecasts are evaluated by their percentage point deviation from the Hodrick-Prescott trend of aggregate EUKLEMS tfp-growth over the forecast horizon.

The differences in timeliness and availability of micro and macro data motivate our the second experiment. Microdata are published at least a year later than macro data. If our components have any forecasting power then lags contain useful information. So the question here is whether lagged micro-aggregated information has marginal value over the concurrent published aggregate variables. Technically, we ran a horse-race between a 'clean' macro-VAR and a 'mixed' micro-macro VAR. That is, we first estimated a simple VAR-model on only aggregate variables. This is basically what we could do if we were in the shoes of the macro-practitioner. Second, we used lagged values of micro-aggregated component values as predetermined variables in the VARs. As a third experiment, we also forecasted aggregate productivity using industry-level subaggregates from the EUKLEMS database.

3.2 Decompositions

3.2.1 Identities

The first dynamic decomposition, laid out by Baily et al.2001, is the following:

$$\Delta\tau_t = \sum_{i\in C}\widetilde{\phi}_i\Delta\tau_{it}+\sum_{i\in C}\Delta\phi_i(\widetilde{\tau}_i-\widetilde{\tau})+\sum_{i\in E}\phi_{it}(\tau_{it}-\widetilde{\tau})-\sum_{i\in X}\phi_{it-1}(\tau_{it-1}-\widetilde{\tau}), \quad (5)$$

where C is the set of continuers, E is the set of entrants and X is the set of exiters in time t. The terms in the previous equation are often called within-, between- and net-entry-terms, respectively.

In the second decomposition, we combine a static decomposition with equation 5. The static equation was introduced by Olley and Pakes1996. We

[11] AR(1) or AR(2) depending on Akaike and Schwartz Information Criteria. We estimated the timeseries models on 17 observations (years 1978-1995), so kept it simple to preserve degrees of freedom.

lay out some notation to help further discussion. Let

$$\begin{aligned}\overline{a}_t &= \frac{1}{N_t}\sum_{N_t} a_{it} \text{ be a cross section average,}\\ \widetilde{a} &= \frac{a_t + a_{t-1}}{2} \text{ denote a time-average}\end{aligned}$$

and the static decomposition be

$$\tau_t = \overline{\tau}_t + \sum_{N_t}(\phi_{it} - \overline{\phi}_t)(\tau_{it} - \overline{\tau}_t), \tag{6}$$

where the term $\sum_{N_t}(\phi_{it} - \overline{\phi}_t)(\tau_{it} - \overline{\tau}_t)$ is often called the covariance-term. This decomposition informs us about characteristics of τ's cross-section distribution. When the covariance-term is high, the weighted average is much higher than the simple average. It also implies that the change in aggregate productivity is the sum of the change in the unweighted average productivity and change in the covariance-term. Differencing 6 yields:

$$\Delta\tau_t = \Delta\overline{\tau}_t + \sum_{N_t}(\phi_{it} - \overline{\phi}_t)(\tau_{it} - \overline{\tau}_t) - \sum_{N_{t-1}}(\phi_{it-1} - \overline{\phi}_{t-1})(\tau_{it-1} - \overline{\tau}_{t-1}), \tag{7}$$

ie. aggregate productivity change is sum of the change in the unweighted average productivity and change in the covariance term. Entry and exit are implicity accounted for here as the cross section sums run from $i = 1...N_t$ and $i = 1...N_{t-1}$ respectively. Hence, the weights are such that $\sum_{N_t}\phi_{it} = 1$ and $\sum_{N_{t-1}}\phi_{it-1}$. Now express $\widetilde{\tau}$ using 6:

$$\begin{aligned}\widetilde{\tau} &= \frac{\tau_t + \tau_{t-1}}{2} =\\ &= \frac{1}{2}\left[\overline{\tau}_t + \overline{\tau}_{t-1} + \sum_{N_t}(\phi_{it} - \overline{\phi}_t)(\tau_{it} - \overline{\tau}_t) + \sum_{N_{t-1}}(\phi_{it-1} - \overline{\phi}_{t-1})(\tau_{it-1} - \overline{\tau}_{t-1})\right]\\ &= \frac{\overline{\tau}_t + \overline{\tau}_{t-1}}{2} + \frac{\sum_{N_t}(\phi_{it} - \overline{\phi}_t)(\tau_{it} - \overline{\tau}_t) + \sum_{N_{t-1}}(\phi_{it-1} - \overline{\phi}_{t-1})(\tau_{it-1} - \overline{\tau}_{t-1})}{2}\\ &= \widetilde{[\overline{\tau}_t]} + \widetilde{cov}(\phi, \tau),\end{aligned} \tag{8}$$

with

$$\begin{aligned}
\widetilde{[\overline{\tau}_t]} &= \frac{\overline{\tau}_t + \overline{\tau}_{t-1}}{2}, \text{ and} \\
\widetilde{cov}(\phi,\tau) &= \frac{1}{2}\left[\sum_{N_t}(\phi_{it}-\overline{\phi}_t)(\tau_{it}-\overline{\tau}_t) + \sum_{N_{t-1}}(\phi_{it-1}-\overline{\phi}_{t-1})(\tau_{it-1}-\overline{\tau}_{t-1})\right]
\end{aligned}$$

$\widetilde{cov}(\phi,\tau)$ is going to reveal information about the cross section characteristics of the distribution. Plugging these terms into 5 yields:

$$\begin{aligned}
\Delta\tau_t &= \sum_C \widetilde{\phi}_i \Delta\tau_{it} + \sum_C \Delta\phi_i(\widetilde{\tau}_i - (\widetilde{[\overline{\tau}_t]} + \widetilde{cov}(\phi,\tau))) \\
&\quad + \sum_E \phi_{it}(\tau_{it} - (\widetilde{[\overline{\tau}_t]} + \widetilde{cov}(\phi,\tau))) - \sum_X \phi_{it-1}(\tau_{it-1} - (\widetilde{[\overline{\tau}_t]} + \widetilde{cov}(\phi,\tau)) \\
&= \sum_C \widetilde{\phi}_i \Delta\tau_{it} + \sum_C \Delta\phi_i(\widetilde{\tau}_i - \widetilde{[\overline{\tau}_t]}) - \widetilde{cov}(\phi,\tau)\sum_C \Delta\phi_i \\
&\quad + \sum_E \phi_{it}(\tau_{it} - \widetilde{[\overline{\tau}_t]}) - \sum_X \phi_{it-1}(\tau_{it-1} - \widetilde{[\overline{\tau}_t]}) \\
&\quad - \widetilde{cov}(\phi,\tau))\left[\sum_E \phi_{it} - \sum_X \phi_{it-1}\right]
\end{aligned}$$

This last equation is the basis for the statistical forecasting model. Note:

$$\begin{aligned}
&-\widetilde{cov}(\phi,\tau)\sum_C \Delta\phi_i - \widetilde{cov}(\phi,\tau))\left[\sum_E \phi_{it} - \sum_X \phi_{it-1}\right] \\
&= -\widetilde{cov}(\phi,\tau)\left[\sum_C(\phi_{it}-\phi_{it-1}) + \sum_E \phi_{it} - \sum_X \phi_{it-1}\right] \\
&= -\widetilde{cov}(\phi,\tau)\left[\sum_C \phi_{it} + \sum_E \phi_{it} - \sum_C \phi_{it-1} - \sum_X \phi_{it-1}\right] \\
&= -\widetilde{cov}(\phi,\tau)\left[\sum_{N_t} \phi_{it} - \sum_{N_{t-1}} \phi_{it-1}\right] = 0
\end{aligned}$$

3.2.2 Working equation

We want to keep the within- and net-entry-terms, but we want to introduce the covariance-term from OP to incorporate information about the cross section distribution. Rewrite the between-term in 5 using 8 as

$$\begin{aligned}\sum_C \Delta\phi_i(\widetilde{\tau}_i - \widetilde{\tau}) &= \sum_C \Delta\phi_i\widetilde{\tau}_i - \sum_C \Delta\phi_i\widetilde{\tau} \\ &= \sum_C \Delta\phi_i\widetilde{\tau}_i - \sum_C \Delta\phi_i \left(\widetilde{[\overline{\tau}_t]} + \widetilde{cov}(\phi, \tau)\right).\end{aligned}$$

Then 5 looks like

$$\begin{aligned}\Delta\tau_t &= \\ &\sum_C \widetilde{\phi}_i \Delta\tau_{it} + \sum_C \Delta\phi_i(\widetilde{\tau}_i - \widetilde{[\overline{\tau}_t]}) - \sum_C \Delta\phi_i\widetilde{cov}(\phi, \tau) \\ &+ \sum_E \phi_{it}(\tau_{it} - \widetilde{\tau}) - \sum_X \phi_{it-1}(\tau_{it-1} - \widetilde{\tau}). \qquad (9)\end{aligned}$$

The ϕ_{it} weights were defined at the beginning such that $\sum_{N_t} \phi_{it} = 1$, and hence $\sum_C \Delta\phi_i$ is not necessarily zero (it would be zero had we defined $\sum_C \phi_{it} = 1$). Therefore the covariance term is not killed by $\sum_C \Delta\phi_i$.

The between-term is now different: it sums up the share-changes weighted by a demeaned variable, but the weights are deviations from the time-average of the unweighted cross-section averages. The covariance-term is shown separately.

Push and pull effects Equation 9 shows a fully statistical decomposition. We put some more structure on firm behavior in the second approach, which intends to gauge two effects: a "pull" and a "push" effect. The approach essentially pins down firms' position relative to the frontier. The pull-equation says that individual productivity growth depends positively on the distance from the frontier, in other words, firms further away are pulled more strongly towards it as technology spreads out:

$$\Delta\tau_t = \beta(\tau_t^F - \tau_{it}) + \eta_{it}, \qquad (10)$$

where τ_t^F is frontier productivity and η_{it} is an autonomous or noise component. A case can be made for both $\beta < 0$ and $\beta > 0$. In the first case firms closer to the frontier can be thought to have greater absorption capacity (greater human capital, which is unmeasured, and hence it shows up as higher productivity), so they can take on new technologies and grow faster. In the $\beta > 0$ case less productive firms tend to grow faster. For instance, these firms can be though as being new and smaller entrants with relatively high growth potential.[12]

Frontier productivity (τ_t^F) for each industry was calculated as the average of the top decile from a truncated distribution. We got rid of the top and bottom 1% of the observations to kill the possibly erratic effects on aggregates.[13] We took a moving average of $\tau_{it}^m = \frac{\tau_{it}+\tau_{it-1}}{2}$ to further guard against extreme observations. The resulting truncated distribution was then used to calculate the average of the top decile.

The push equation is assumed to encapsulate the market selection mechanism whereby less efficient firms are crowded out of the market. Assume the following specification:

$$\phi_{it} = \gamma_0 + \gamma_1(\tau_{it} - \overline{\tau}) + \varepsilon_{it}. \tag{11}$$

This equation reads: higher-than-average productivity firms end up gaining market share ($\gamma_1 > 0$), and lower-than-average productivity firms lose market share. The use of equations 10 and 11 is that once we estimated parameters β, γ_0 and γ_1, the fitted values $\widehat{\Delta\tau_t}$ and $\widehat{\phi_{it}}$ from these simple models inform about the real[14] catchup components and shares. If they have some relevant content in our dataset, the noise-free fitted values should have forecasting power in our decompositions. So technically, what we do here is to use the fitted values from these regressions in decompositions 5 and 9.

[12]The literature calls the $\beta > 0$ case - when less productive firms grow faster - as β-convergence. The concept was introduced by Barro and i Martin1991.

[13]I ignored the top and bottom 1% of the productivity distribution because it turned out that frontier time-series exhibited unplausibly large time-variation if these observations were included.

[14]Real in the sense of the simple econometric models above.

Estimation of the auxiliary models The distance variables in both 10 and 11 are endogenous. For 10, the endogeneity problem is immediate. As for 11, a case can be made that a firm gains market share exactly because its productiviy increased in the wake of a postive productivity shock. One way to get around it is to apply an IV estimator. The population moment condition we need for 10 is

$$E[z_i \left(\Delta\tau_i - \widehat{\beta}_{IV}(\tau_t^F - \tau_{it}) \right)] = 0,$$

and the simple Anderson-Hsiao-type IV estimator is given by

$$\widehat{\beta}_{IV} = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T} z_{it}\Delta\tau_{it}}{\sum_{i=1}^{N}\sum_{t=2}^{T} z_{it}(\tau_t^F - \tau_{it})} \text{ with } z_{it} = (\tau_{t-2}^F - \tau_{it-2}).$$

The instruments can also be Δz_{it} but that results in the loss of more observations. As for the push-equation (11), rewriting $\phi_{it} = x\gamma' + \varepsilon_{it}$ where $\gamma = [\gamma_0, \gamma_1]$ and $x = [1, \tau_{it} - \overline{\tau}]$. The conventional IV-estimator again is

$$\begin{aligned}
\widehat{\gamma}_{IV} &= \frac{\sum_{i=1}^{N} z_i'\phi_i}{\sum_{i=1}^{N} z_i' x_i} = \left(Z'X\right)^{-1} Z'\Phi \text{ with} \\
z_i &= [(\tau_{i1} - \overline{\tau}_{i1}), ..., (\tau_{iT-1} - \overline{\tau}_{T-1})]', \\
\phi_i &= [\phi_{i2}, ..., \phi_{iT}]' \text{ and} \\
Z &= [z_1'...z_N']', \Phi = [\phi_1'...\phi_N']'.
\end{aligned}$$

3.3 Forecast results

In every experiment, we compare the relative performance of the micro-aggregated models and their macro-alternatives. The common benchmark for all forecasting models is always the average (over the forecast horizon) growth rate of the Hodrick-Prescott-trend of the published aggregate. The

HP-trend is not observed over the forecast period (out-of-sample) because the filtering procedure does not allow direct forecasting so we could not use the extrapolation of the HP-trend as a macro-alternative.

In the first experiment, the macro-alternative is taken to be the simple trend growth. The second experiment uses VAR-forecasts as the most obvious macro-alternatives. Tables 6 and 7 show detailed in-sample and out-of sample forecast performance, respectively. The first column of each table identifies whether direct measures or fitted values were used for $\Delta\tau_{it}$ and ϕ_{it} in the decompositions. The second column shows the names of forecasted terms in the decompositions. The abbreviations *with*, *betw* and *nex* stand for within-, between- and net-entry-components. Columns 3-4 contain the forecasted series' average growth rate in the two decompositions (equations 5 and 9) for value-added-based productivity. Columns 5-6 show the same for revenue-based productivity.

The lower panel of tables 6 and 7 show benchmarks against which we could judge forecast performance. *Macro tfp actual* is the average growth rate of the observed EUKLEMS-aggregate over the forecast horizon. *Macro tfp trend* is the average of Hodrick-Prescott filtered-trend growth rates (observed). *Macro tfp trend forecast* is the average of the forecasted yearly trend growth rates from a regression of the aggregate level on a time-trend and a constant. The estimation and forecast-periods are always indicated in the captions of the tables.

To keep things tractable, we comment on tables 6 and 7 only in passing[15] and turn to tables 8 and 9 for more direct conlcusions. The general patterns from tables 6 and 7 are: observed and forecasted aggregate rates are smaller for output-based producitivity both in- and out-of-sample; the growth rate of the actual tends to be larger than that of the trend series (the trend smoothes out large swings).

Tables 8 and 9 show forecast performance relative to the average of the observed HP-trend growth. The first entry of table 8 says that the a value-added-based forecast (using measured [not fitted] producitivity and observed

[15] These tables are not reported for other forecasting exercises.

[not fitted] shares) from the micro aggregate is 0.9% lower than the HP-trend over 2000-2004. All the other entries are to read analogously. It is clear that neither of the micro-aggregated forecasts could beat the simple trend growth specification in-sample because it was very close to the HP-trend growth (which we treat as the target).[16] It holds irrespective of decompositions or models and whether fitted/observed values were used.

Results are more promising out-of-sample for the value-added-model, which is of main interest here. We see from column 2 in table 9 that both value-added-based decompositions perform better than the macro-alternative. It is remarkable that the micro-aggregated model for value added would have forecasted the productivity boom of the second half of the nineties better than the simple macro-alternative. The -1.1 percentage point deviation of the macro alternative suggests that the macro forecast would have missed the productivity acceleration. The output model performs poorly relative to the value added model (see also section 2.2).

The results of experiment 2 are summarized in table 10 for the value added model. Since we estimated VARs on a small sample, we clung to parsimonious specifications just like at single equation techniques in the previous experiment. In short, we estimated two VARs with maximum lag order of 1 or 2. The first VAR is the 'clean' macro-VAR containing only EUKLEMS-aggregates[17] as endogenous variables. The second VAR is the 'mixed' micro-macro VAR containing both predetermined[18] and endogenous[19] variables (see the addenda of tables 10 and 11 for more details). The last row of each table contains the forecast performance of the macro-alternative. Table 10's 1.53 in the last row means that the system's forecasted tfp growth series is on average 1.53 percentage point higher over 1998-1999-2000 than the average growth rate of the HP-filtered aggregate. The first five rows contain similar deviations for the 'mixed' micro-aggregated-macro VARs. It is apparent that

[16] Although decomposition 1 with fitted shares and productivity gives the same forecast for the output model (next-to-last entry of column 4 in table 8).

[17] Yearly growth rates apital and labor services, value added/output, and tfp.

[18] Components in equations 5 and 9.

[19] Yearly growth rates apital and labor services, value added/output, and tfp.

'mixed' VARs featuring the components of equation 5 (dc1) do better than the 'clean' macro-VAR (first five entries of column 2 and 4 and the last entry of the column). Although results for equation 9 (dc2) seem to be sensitive to the VAR-specification, the message of this table is simple: it is beneficial to use lagged micro data *and* concurrent macro-aggregates. Results also suggest that it is to the purpose to use more lags for the macro than for the micro-aggregated variables. These results did not change when the aggregate VAR was replaced by another vector autoregression containing productivity subaggregates within overall manufacturing. Another way of saying this is that the 'mixed' micro-macro VAR beat both aggregate forecasts. The table for the latter aggregate VAR is not shown, we only note here that the equivalent of the entry 1.53 in the last row of table 10 is -1.6.

Results are 'poor' for the output model for reasons outlined beforehand (table 11).

4 Conclusion

The aim of this paper was to build forecasts based on firm-level productivity estimates. We line up along the productivity literature in that we build micro-level productivity measures but our question was *not* how to measure productivity at the firm level but rather how to apply these estimates in forecasting. To our knowledge, our study is a novel attempt to connect micro and macro level analysis whereby micro-level productivity estimates are used to build aggregates and forecasts.

We carried out several forecasting experiment for different specifications and models. Although there are some remaining questions, our results suggest that components from productivity decompositions improve aggregate tfp forecasts.

References

Ackerberg, D. A. & Caves, K. (2004). Structural identification of production functions - an application to the timing of input choice. mimeo.

Baily, M. N., Bartelsman, E. J., & Haltiwanger, J. (2001). Labor productivity: Structural change and cyclical dynamics. *The Review of Economics and Statistics*, 83(3):420–433.

Barro, R. & i Martin, X. S. (1991). Convergence across states and regions. *Brookings Papers on Economic Activity*, 1:107–182.

Bartelsman, E. J. & Doms, M. (2000). Understanding productivity: lessons from longitudinal microdata. *Journal of Economic Literature*, 38(3):569–594.

Bartelsman, E. J., Haltiwanger, J., & Scarpetta, S. (2005). Measuring and analyzing coss-country differences in firm dynamics. draft March 2005, Presented at NBER/CRIW conference, Washington DC.

Corrado, C., Beaulieu, J., Bartelsman, E. J., & Lengermann, P. (2006). Modeling aggregate labor productivity and a disaggregate level: New results for us sectors and industries. mimeo.

Domar, E. D. (1961). On the measurement of technological cha. *The Economic Journal*, 71(284):709–729.

Fagan, G., Henry, J., & Mestre, R. (2001). An area-wide model (awm) for the euro area. *ECB Working Paper*, 2001(42).

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

Javocik, B. S. (2004). Does foreign direct investment increase the productivity of domestic firms? in search of spillovers through backward linkages. *The American Economic Review*, 98(3):605–627.

Kátay, G. & Wolf, Z. (2008). Driving factors of growth in hungary - a decomposition exercise. Working paper, Magyar Nemzeti Bank. forthcoming.

Levinsohn, J. & Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies*, 70(2):317–341.

Oliner, S. D., Sichel, D. E., & Stiroh, K. J. (2007). Explaining a productive decade. *Finance and Economics Discussion Series, Board of Governors of the Federal Reserve System*, 2007(63).

Olley, G. S. & Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297.

O'Mahony, M., Timmer, M. P., & Ark, B. V. (2007). Euklems growth and productivity accounts: Overview november 2007 release. mimeo.

Petrin, A. & Levinsohn, J. (2008). Measuring aggregate productivity growth using plant-level data. mimeo.

5 Appendix

5.1 The data

5.1.1 CBS data

General description The microlevel database consists of a large panel of Dutch manufacturing firms. The database is under close surveillance by the Centrum voor Beleidsstatistiek[20]. The time dimension of the panel spans between 1978 and 2004.

To compile our working data, we first established longitudnial links in the database because the CvB changed firm identifiers as of 1993. We used correspondences between firm identifiers included in the underlying datafiles.

[20] Dutch Statistics Office

There was no documentation about such aspects of firm dynamics as mergers, acquisitions and split-ups so we could not control for these transformations.

The industry classification system also changed in 1993. Since both the 1974-based ('old') and the 1993-based ('new') industry codes were available in 1990-1992, we generated a correspondence between the two industry classifications. We could then use this correspondence to generate deflators that originated from the EUKLEMS dataset, which in turn is based on the new industry classification.

Table 1 explains names of industry codes. In order not to lose observations and maintain a reasonable size, we merged industries 23 and 24.

Table 2 describes how many firms were used in the decompositions, the last columns of tables 3 and 4 show the how many observations were used in tfp-estimation. The last columns in tables 3 and 4 show firm-year observations whereas table 2 treats each firm as a single observation.

Variables The dataset did not have observed capital so we used depreciation as a proxy for the capital stock. We deflated nominal depreciation using an industry-specific implicit capital deflator that was calculated from EUKLEMS data. The other firm-specific micro-variables were: number of employees, output, value added, input materials, energy and payroll.

Outliers We filtered outliers using the interquartile-range method. This method is a standard way of dealing with skewed distributions. For tfp-estimation, we considered an observation to be outlier if either of the following was true:

$$\begin{aligned} x_o &< q25 - 1.5 * IQR \\ x_o &> q75 + 1.5 * IQR, \end{aligned}$$

where $x_o \equiv \{K/Y; M/Y; WEXP/L\}$ and K denotes capital, M denotes input material purchases, $WEXP$ denotes wage expenditures, Y denotes either value added or output depending on which model we used. $q25$ and

$q75$ stand for the lower and upper quartiles of the distribution of the variables in the set x_o, and $IQR \equiv q75 - q25$.

For aggregation, we used slightly different filtering. Since depreciation is not observed in the first six years of the panel, we had to backcast individual productivity growth rates in 1978-1983. The backcasting procedure was based on the (geometric) average of nonoutlier valueadded/revenue labor-productivity growth. Using these "imputed" productivity numbers we carried out a secondary outlier filtering in order to get rid of implausibly large or small productivity-observations. The logic of this secondary filtering was the same as above with the exception that we deployed energy instead of depreciation.

5.1.2 The EUKLEMS data

History and general description The EUKLEMS "... *project aims to create a database on measures of economic growth, productivity, employment creation, capital formation and technological change at the industry level for all European Union member states from 1970 onwards. This work will provide an important input to policy evaluation, in particular for the assessment of the goals concerning competitiveness and economic growth potential as established by the Lisbon and Barcelona summit goals. The database should facilitate the sustainable production of high quality statistics using the methodologies of national accounts and input-output analysis. The input measures will include various categories of capital, labour, energy, material and service inputs. Productivity measures will be developed, in particular with growth accounting techniques. Several measures on knowledge creation will also be constructed. Substantial methodological and data research on these measures will be carried out to improve international comparability. There will be ample attention for the development of a flexible database structure, and for the progressive implementation of the database in official statistics over the course of the project. The database will be used for analytical and policy-related purposes, in particular by studying the relationship between skill formation, technological progress and innovation on the one hand, and pro-*

ductivity, on the other. To facilitate this type of analysis a link will also be sought with existing micro (firm level) databases. The balance in academic, statistical and policy input in this project is realised by the participation of 15 organisations from across the EU, representing a mix of academic institutions and national economic policy research institutes and with the support from various statistical offices and the OECD."[21]

Structure There have been two EUKLEMS releases so far. We used the one that was published in March 2008. The files contain 62 variables, including basic data and growth accounting variables. The datafiles are structured to follow an industry classification list which corresponds with the NACE[22] list.

Variables For a detailed description of the variables, see **?**. This project used growth accounting variables in forecasting exercises and deflators in the nonparametric identification procedure. To be more specific, the following variables were used in forecasting experiments.
GO_QI: Gross output, volume indices, 1995 = 100,
VA_QI: Gross value added, volume indices, 1995 = 100,
LAB_QI: Labour services, volume indices, 1995 = 100,
CAP_QI: Capital services, volume indices, 1995 = 100,
TFPva_I: TFP (value added based), 1995=100,
TFPgo_I: TFP (gross output based) growth, 1995=100,
GO: Gross output at current basic prices (in millions of local currency),
VA: Gross value added at current basic prices (in millions of local currency).

The following were used for firm-level calculations at the first stage of the project:
GO_P: Gross output, price indices, 1995 = 100,
VA_P: Gross value added, price indices, 1995 = 100,

[21] http://www.euklems.net
[22] NACE – Classification of Economic Activities in the European Community.

II_P: Intermediate inputs, price indices, 1995 = 100,

CAP: Capital services (in millions of local currency),

CAP_QI: Capital services, volume indices, 1995 = 100,

the last two were used to calculate implicit capital deflators for every industry.

Tables and figures

Table 1: Key to industry codes

Manufacturing classification	
industry code	
Consumer manufacturing	
15t16	Food products, beverages and tobacco
17t19	Textiles, textile products, leather and footwear
36t37	Manufacturing nec; recycling
Intermediate manufacturing	
20	Wood and products of wood and cork
21t22	Pulp, paper, paper products, printing and publishing
23	Coke, refined petroleum products and nuclear fuel
24	Chemicals and chemical products
25	Rubber and plastics products
26	Other non-metallic mineral products
27t28	Basic metals and fabricated metal products
Investment goods, excluding hightech	
29	Machinery, nec
34t35	Transport equipment
Electrical machinery and communication services	
30t33	Electrical and optical equipment

Table 2: Number of firms in the analysis

industry code	original	va model	output model
15t16	5994	4284	4421
17t19	3823	2503	2429
20	1321	1040	1038
21t22	5908	4064	4224
24	1585	1088	1132
25	1713	1307	1321
26	2076	1395	1472
27t28	7617	5117	5246
29	4804	3358	3367
30t33	3296	2140	2197
34t35	2291	1504	1534
36t37	3743	2205	2265
overall manufacturing	44171	30005	30646

Table 3: Estimation results - Value added model

industry code	L	K	M	ctrs	seL	seK	nobs
15t16	0.52	0.22	.	581.607	0.007	0.014	19745
17t19	0.54	0.23	.	390.991	0.008	0.011	9457
20	0.58	0.12	.	431.116	0.017	0.015	4624
21t22	0.60	0.13	.	660.952	0.009	0.006	20196
24	0.44	0.23	.	101.387	0.016	0.023	6390
25	0.53	0.24	.	84.735	0.019	0.015	6800
26	0.52	0.21	.	191.860	0.013	0.016	7094
27t28	0.65	0.12	.	1056.116	0.008	0.008	24428
29	0.62	0.12	.	458.341	0.012	0.010	18231
30t33	0.65	0.16	.	211.017	0.013	0.010	8008
34t35	0.72	0.06	.	183.449	0.014	0.010	6861
36t37	0.59	0.13	.	322.399	0.013	0.012	9033

Table 4: Estimation results - Output model

industry code	L	K	M	ctrs	seL	seK	nobs
15t16	0.25	0.12	0.65	2.6439	0.013	0.019	19745
17t19	0.40	0.16	0.42	2.9865	0.019	0.022	9457
20	0.31	0.01	0.66	0.0296	0.012	0.007	4624
21t22	0.36	0.07	0.56	0.0657	0.008	0.033	20196
24	0.23	0.13	0.59	2.0937	0.017	0.153	6390
25	0.33	0.02	0.59	1.8747	0.011	0.047	6800
26	0.34	0.02	0.69	0.0368	0.010	0.152	7094
27t28	0.45	0.01	0.54	0.0002	0.008	0.003	24428
29	0.42	0.06	0.53	0.4290	0.007	0.018	18231
30t33	0.47	0.05	0.63	2.8980	0.013	0.036	8008
34t35	0.44	0.01	0.55	0.0194	0.015	0.000	6861
36t37	0.38	0.1	0.37	6.7575	0.013	0.020	9033

Table 5: Tfp-growth in overall manufacturing in different countries (average percent per annum rates over period indicated in first column)

	USA	UK	GER	FRA	IRL	NED
1977-1982	-0.010	0.002	0.005	#N/A	#N/A	0.004
1983-1987	0.025	0.026	0.005	0.002	#N/A	0.014
1988-1993	0.007	0.026	0.013	0.007	#N/A	0.006
1994-1999	0.022	-0.010	0.008	0.029	0.044	0.016
2000-2005	0.035	0.026	0.015	0.015	-0.006	0.016

Table 6: Forecast results: average tfp-growth rates for overall EUKLEMS manufacturing (percent per annum, sample: 1978-2004, forecast period: 2000-2004).

(2000-2004)

structure	component	model			
		value added		output	
		dc1	**dc2**	**dc1**	**dc2**
va/q shares	with	1.2	1.2	0.2	0.2
	betw	0.0	0.1	0.0	0.0
	nex	-0.3	-0.3	0.0	-0.1
	aggr	***1.0***	***1.0***	***0.2***	***1.5***
input shares	with	1.1	1.1	0.2	1.4
	betw	0.0	0.1	0.0	0.0
	nex	-0.5	0.0	0.0	0.0
	aggr	***0.6***	***1.2***	***0.3***	***1.4***
fitted va/q shares & tau	with	3.2	3.2	2.4	9.2
	betw	0.4	-0.3	0.0	-0.2
	nex	-2.4	-2.4	-0.8	0.0
	aggr	***1.2***	***0.6***	***1.6***	***9.0***
fitted input shares & tau	with	2.0	2.0	1.1	3.9
	betw	0.3	0.1	0.0	0.0
	nex	-1.5	-1.5	-0.6	0.0
	aggr	***0.9***	***0.6***	***0.5***	***3.9***
average growth rates of benchmarks:					
Macro tfp actual*		***2.3***		***0.7***	
Macro tfp trend**		***1.9***		***0.6***	
Macro tfp trend forecast***		***1.7***		***0.5***	

**Actual tfp growth rates from macro data*

***HP-trend tfp growth from macro data*

****Regression based tfp trend growth from macro data*

dc1: equation 5; dc2: equation 9, actual: average of observed growth rates; hp-trend growth: average of Hodrick-Prescott filtered-trend growth rates (observed); simple trend growth: average of trend growth rates from a regression on a time-trend and a constant (forecasted).

Table 7: Forecast results: average tfp-growth rates for overall EUKLEMS manufacturing (percent per annum, sample: 1978-1995, forecast period: 1996-2000).

(1996-2000)

structure	component	model			
		value added		output	
		dc1	**dc2**	**dc1**	**dc2**
va/q shares	with	1.8	1.8	1.6	1.6
	betw	0.0	0.0	0.0	0.0
	nex	0.2	0.2	-0.1	-0.1
	aggr	***1.9***	***2.0***	***1.5***	***1.5***
input shares	with	1.6	1.6	1.4	1.4
	betw	0.0	0.0	0.0	0.0
	nex	0.0	0.0	-0.2	0.0
	aggr	***1.6***	***1.6***	***1.2***	***1.4***
fitted va/q shares & tau	with	2.2	2.2	9.2	9.2
	betw	0.1	0.1	0.1	-0.2
	nex	0.0	0.0	0.0	0.0
	aggr	***2.3***	***2.3***	***9.2***	***9.0***
fitted input shares & tau	with	1.4	1.4	3.9	3.9
	betw	0.1	0.0	0.0	0.0
	nex	0.0	0.0	0.0	0.0
	aggr	***1.5***	***1.4***	***3.9***	***3.9***
average growth rates of benchmarks:					
Macro tfp actual*		***2.1***		***0.6***	
Macro tfp trend**		***1.9***		***0.6***	
Macro tfp trend forecast***		***0.8***		***0.2***	

**Actual tfp growth rates from macro data*

***HP-trend tfp growth from macro data*

****Regression based tfp trend growth from macro data*

dc1: equation 5; dc2: equation 9, actual: average of observed growth rates; hp-trend growth: average of Hodrick-Prescott filtered-trend growth rates (observed); simple trend growth: average of trend growth rates from a regression on a time-trend and a constant (forecasted).

Table 8: In-sample relative forecast performance, (percentage point deviation from HP-trend of EUKLEMS manufacturing aggregate, sample: 1978-2004, forecast period: 2000-2004)

structure	value added		output	
	dc1	**dc2**	**dc1**	**dc2**
va/q shares	-0.9	-0.8	-0.4	0.9
input shares	-1.3	-0.7	-0.3	0.8
fitted va/q shares & tau	-0.6	-1.3	1.0	8.4
fitted input shares & tau	-1.0	-1.2	-0.1	3.3
Macro tfp trend forecast*	***-0.2***	***-0.2***	***-0.1***	***-0.1***

**Regression based tfp trend growth from macro data*

dc1: equation 5; dc2: equation 9; simple trend growth: average of trend growth rates from a regression on a time-trend and a constant (forecasted).

Table 9: Out-of-sample relative forecast performance, (percentage point deviation from HP-trend of EUKLEMS manufacturing aggregate, sample: 1978-1995, forecast period: 1996-2000)

structure	value added		output	
	dc1	**dc2**	**dc1**	**dc2**
va/q shares	0.1	0.1	1.0	0.9
input shares	-0.2	-0.2	0.6	0.9
fitted va/q shares & tau	0.4	0.4	8.7	8.4
fitted input shares & tau	-0.4	-0.5	3.3	3.3
Macro tfp trend forecast*	***-1.1***	***-1.1***	***-0.3***	***-0.3***

**Regression based tfp trend growth from macro data*

dc1: equation 5; dc2: equation 9; simple trend growth: average of trend growth rates from a regression on a time-trend and a constant (forecasted).

Table 10: Out-of-sample relative forecast performance of value added model, experiment 2 (percentage point deviation from HP-trend of EUKLEMS manufacturing aggregate, sample: 1978-1997, forecast period: 1998-2000)

structure	VAR(1,2)*		VAR(2,1)**	
VAR - micro & macro	**dc1**	**dc2**	**dc1**	**dc2**
va/q shares	0.25	9.17	0.81	-0.63
input shares	0.42	9.40	0.86	-0.73
fitted va/q shares & tau	0.36	9.45	0.81	-0.64
fitted input shares & tau	0.31	9.43	0.84	-0.66
VAR - macro	***1.53***	***1.53***	***1.53***	***1.53***

** VAR(1,2): 1st lag of endogenous variables (growth rates of aggregate capital services, labor sevices, tfp, value added), 1st and 2nd lags of exogenous variables. Exogenous variables are lagged values of components in the micro-based decompositions (ie. X(t-1) and X(t-2)).*

***VAR(2,1): 1st and 2nd lags of endogenous variables (growth rates of aggregate capital services, labor sevices, tfp, value added), 1st lag of exogenous variables. Exogenous variables are lagged values of components in the decompositions (ie. X(t-1)).*

dc1: equation 5; dc2: equation 9; dynamic solutions of the VARs were calculated using the Gauss-Seidel algorithm.

Table 11: Out-of-sample relative forecast performance of output model, experiment 2 (percentage point deviation from HP-trend of EUKLEMS manufacturing aggregate, sample: 1978-1997, forecast period: 1998-2000)

structure	VAR(1,2)*		VAR(2,1)**	
VAR - micro & macro	**dc1**	**dc2**	**dc1**	**dc2**
va/q shares	0.211	-0.283	0.406	0.393
input shares	0.226	-0.276	0.406	0.398
fitted va/q shares & tau	0.205	-0.259	0.409	0.389
fitted input shares & tau	0.212	-0.250	0.402	0.386
VAR - macro	***0.028***	***0.028***	***0.028***	***0.028***

** VAR(1,2): 1st lag of endogenous variables (growth rates of aggregate capital services, labor sevices, tfp, output), 1st and 2nd lags of exogenous variables. Exogenous variables are lagged values of components in the micro-based decompositions (ie. X(t-1) and X(t-2)).*

***VAR(2,1): 1st and 2nd lags of endogenous variables (growth rates of aggregate capital services, labor sevices, tfp, value added), 1st lag of exogenous variables. Exogenous variables are lagged values of components in the decompositions (ie. X(t-1)).*

dc1: equation 5; dc2: equation 9; dynamic solutions of the VARs were calculated using the Gauss-Seidel algorithm.

Figure 1: Tfp-growth in overall manufacturing in different countries.

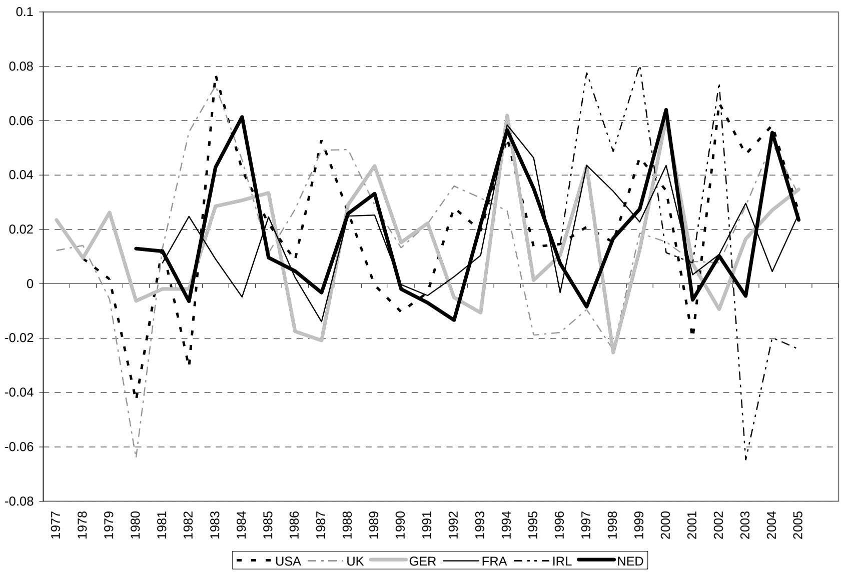

Figure 2: Tfp growth by micro-based and EUKLEMS-based aggregate; overall manufacturing.

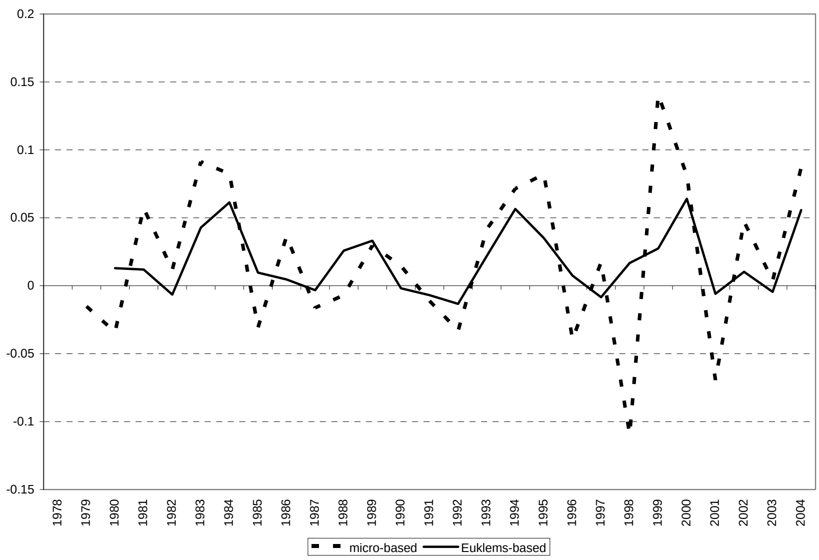

Figure 3: Value added growth by micro-based and EUKLEMS-based aggregates; overall manufacturing (solid lines: micro-based aggregate; dashed line: EUKLEMS aggregate).

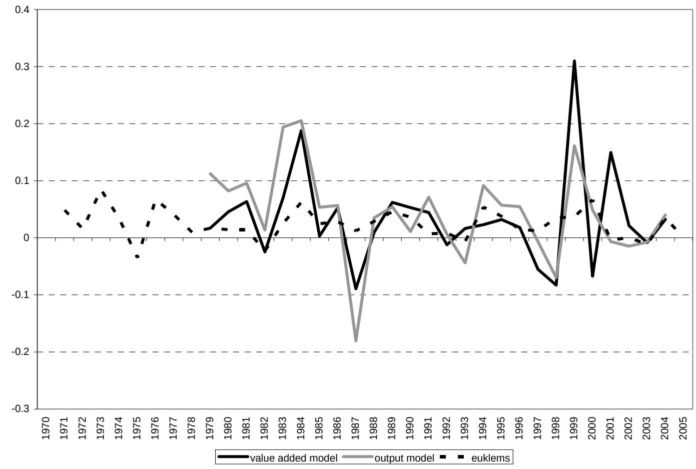

Figure 4: Output growh by micro-based and EUKLEMS-based aggregates; overall manufacturing (solid lines: micro-based aggregate; dashed line: EUKLEMS aggregate).

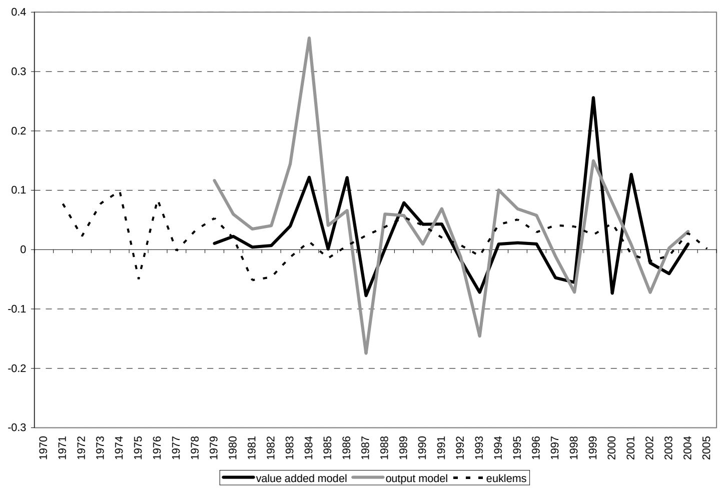

Figure 5: Employment growth by micro-based and EUKLEMS-based aggregates; overall manufacturing (solid lines: micro-based aggregate; dashed line: EUKLEMS aggregate).