Using Census Microdata to Forecast U.S. Aggregate Productivity

Eric J. Bartelsman[a,b], Christopher J. Kurz[c], Zoltan Wolf[a,*†]

[a] Vrije Universiteit Amsterdam, Tinbergen Institute

[b] IZA Bonn

[c] Board of Governors of the Federal Reserve System

July 11, 2011

Abstract

We contribute to the productivity literature by building a bridge between time series analysis of macro-level productivity growth and firm-level productivity studies. We confirm earlier results that information from firm-level data improves univariate and multivariate forecasts of aggregate productivity using data from the U.S. We extend a state-space model to forecast medium term productivity growth by including a role for heterogeneous firms and the evolution of firm-level productivity and size distributions. We find that adding time series that capture the contribution to productivity growth of within firm productivity growth and reallocation for US manufacturing firms improves estimates and forecasts of trend productivity growth for both the manufacturing and the aggregate non-farm business sectors.

keywords: Total factor productivity, firm-level data, aggregation, forecasting, Kalman filter

1 Introduction

Potential output is a key variable in policy functions of central banks owing to its implications for inflationary pressure. Unfortunately, potential output is unobserved. Macro models (e.g. (Fagan *et al.*, 2001) or (Roberts, 2001)) back out trend potential output using the assumption that long-run and short-run determinants of output are different. Such trend-cycle decompositions consider trend output to be governed by the long-run evolution of production factors and productivity. The development of trend inputs – capital, labor hours, (and labor quality) – is governed by conditions derived from a representative-agent long-run growth model. In these forecasting models, trend productivity essentially is a residual, the trend that is left over

*Address for correspondence: VU University Amsterdam, Department of Economics, Room 2E-72, De Boelelaan 1105, 1081 HV Amsterdam The Netherlands. Tel: +31 20 598 6030. Fax +31 20 598 9870. E-mail: z.wolf@vu.nl.

†The views expressed herein are solely the authors' and do not necessarily reflect those of the Board nor the staff of the Federal Reserve System.

after all other trends have been removed. Therefore, improved forecasts of trend total factor productivity (TFP) would be helpful in generating better estimates of potential output.

Total factor productivity growth (TFP) simply can be defined as the difference of the growth rates of output and the weighted growth rates of inputs.[1] At the level of the aggregate economy, one could define TFP as the change in real final output minus the change in primary inputs needed to achieve the change in output. Productivity is hard to measure, with considerable discussion of appropriate theoretical constructs and numerous empirical sources of bias and error.[2] Moreover, even the most carefully constructed productivity indexes exhibit growth rates that fluctuate wildly from year to year.[3] Productivity thus appears to be an elusive target for forecasting exercises.

Our understanding of the underlying sources of aggregate productivity movements also is not as well developed as, e.g. our understanding of the mechanisms underlying firm dynamics or the developments in factor markets. Recent work grounded in empirical observation ((Bartelsman and Doms, 2000), (Bartelsman *et al.*, 2005), and (Foster *et al.*, 1997)) lays the source of productivity change at explicit actions at the firm level, so that aggregate TFP change is the result of within-firm changes in productivity and between-firm movements in market share. In this environment, forecasting aggregate productivity movements may be improved with firm-level measures of activity, productivity, and size.[4] While recent studies such as (Petrin and Levinsohn, 2010), (Petrin *et al.*, 2009) and (Basu *et al.*, 2009) link microeconomic productivity movements to changes at the aggregate level, our goal is to improve forecasts of aggregate productivity for the U.S. by integrating the framework of firm-level dynamics with macro trend-cycle models.

Our paper does not formally model an economy of heterogeneous firms, where their explicit actions and market interactions lead to aggregate productivity movements. In other work, (Wolf, 2011), runs simulations of the (Bloom, 2009) model with heterogeneous firms in an uncertain environment, to show how the productivity components do aid in uncovering the sources of change in aggregate productivity. Intuitively, one can think of three components of aggregate productivity growth: "within," "between," and "net-entry" terms (Baily *et al.*, 1992). First, the 'within' component of productivity growth is a simple weighted average of firm-level growth rates. It is the sole factor behind aggregate productivity dynamics in a neoclassical

[1]The BLS defines MFP (TFP) as a measure of the joint influences on economic growth of technological change, efficiency improvements, returns to scale, reallocation of resources due to shifts in factor inputs across industries, and other factors.

[2]See, (Hulten, 2001) for a summary of the theory behind total factor productivity and some sources of potential biases.

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

[4]For work on estimating micro-level total factor productivity see : (Olley and Pakes, 1996), (Levinsohn and Petrin, 2003) and (Wooldridge, 2009). For work at the aggregation and at the aggregate level see: (Domar, 1961), (Hulten, 1978).

world. The within component can be expected to have forecasting power because it captures common behavior over the business cycle (such as factor hoarding), and steady factors at lower frequencies (such as technology diffusion). Second, the 'between' component is assumed to represent the underlying forces of reallocation of resources across firms. It captures market selection mechanisms whereby more productive firms gain and less productive firms lose market share. Based on the findings of earlier work by (Baily *et al.*, 2001), (Basu and Kimball, 1997) and (Basu *et al.*, 1998) it may be expected to behave quite differently over the business cycle than the within component. Third, the net entry component captures the entry and exit on the margin, as well as the rejuvenation of industries through high-growth startup. A large net entry component implies that entrants' productivity is larger than exiters', possibly indicating new opportunities arising through technological breakthroughs.[5]

In this paper, we expand upon the work of (Bartelsman and Wolf, 2009) who find that information from firm-level data improve univariate and multivariate forecasts of aggregate productivity for Dutch manufacturing firms. We confirm their results using plant-level data from the U.S. Census Bureau to aid in forecasting macro-level productivity. Our main goal is to see whether we can extend the state-space models used to forecast potential output to accommodate the view that aggregate productivity is comprised of within-firm productivity and reallocation terms. More practically, we want to assess whether using the firm-level productivity components actually improve trend productivity forecasts generated by the model. In particular, we adopt and modify (Roberts, 2001) time-varying parameter techniques of estimating trend productivity using a state space framework.[6] We used plant-level data from the Census of Manufactures, the Annual Survey of Manufactures, and the Longitudinal Research Database to create a sample of U.S. manufacturing firms from 1974-1998.[7] The aggregate-level data we use come from the Bureau of Labor Statistics (BLS) Multifactor Productivity Release.

Our results indicate that using microdata for the U.S. manufacturing sector improves univariate and multivariate forecasts of aggregate TFP growth. This implies that there is information content in the micro-aggregated components that assists in forecasting a productivity series, independent of the series itself. Moreover, we find an intuitive and convenient way to extend the state-space forecasting model to accommodate firm-level productivity components. When placing our estimated micro-aggregated components within this framework, we can also improve upon estimates and forecasts of structural productivity.

The remainder of the paper is organized as follows: We start with a description of the data used to estimate plant-level TFP and aggregate productivity components, and of the industry

[5] The Bloom model, as used by Wolf, does not have explicit entry and exit mechanisms.

[6] (Roberts, 2001) provides the foundation for estimating trend productivity at the Board of Governors of the Federal Reserve System. We would like to thank Bruce Fallick, Charles Fleischmann, and John Roberts for providing assistance in explicating the state space estimation procedures.

[7] We use plant-level data to estimate productivity and generate time-series components. For sake of generality we refer to 'firm-level', except when we specifically describe our empirical work with the U.S. data.

and macro data used in the forecasting exercises. Next, we describe briefly our methods to compute plant-level TFP and the productivity decompositions. We then provide the methods and results on using plant-level information to improve univariate and multivariate productivity forecasts. The main contribution of our paper discusses how to use micro-components to estimate structural productivity in a state space model, and presents the results for forecasting manufacturing TFP and private nonfarm business sector productivity. We conclude with a discussion of caveats and thoughts for future work.

2 Data and aggregation

We now describe the plant-level data used to estimate productivity and to calculate the within plant and reallocation contributions to aggregate productivity. Next, we present comparisons of our micro-aggregated measures to published data from the Bureau of Labor Statistics (BLS).

2.1 Data

The plant-level data employed to estimate the productivity measures and calculate micro-aggregated components sources from the Annual Survey of Manufactures (ASM), the Census of Manufactures (CMF), and the Longitudinal Business Database (LBD). The data span 25 years from 1974 to 1998 and contain roughly 45-50 thousand plants a year and about 1.2 million plants overall.

Despite the richness of the Census microdata, there are several drawbacks. The CMF only takes place every five years and the ASM is a sample of manufacturing plants. The ASM sample contains waves of plants that are not continually present within the entire panel. To solve these problems, we utilize ASM weights that allow for industry-level aggregates to be estimated and use the LBD to create our longitudinal links in order to properly account for plant entry and exit.

The ASM provides the majority of the data within this project. The goal of the ASM is to sample from the universe of manufacturing establishments to provide estimates of statistics for all manufacturing establishments with one or more paid employee. The ASM includes roughly 50,000 establishments from the census universe of about 400,000 manufacturing establishments and represents roughly 70 percent of total manufacturing shipments. Within the 50,000 sampled plants, about 20 percent are sampled with certainly and the remaining establishments are selected with a probability proportional to establishment size, i.e., probability weights.[8] The ASM is conducted in the 4 years between the economic census years.[9] The variables of interest from the ASM include plant level identifiers, industry affiliation,[10] measures of capital and

[8] Typically, plants with 250 or more employees are sampled with certainty. In 2004, the ASM sample design changed.

[9] Our panel of interest contains five, five-year ASM sub-panels: 1974-1978, 1979-1983, 1983-1988, 1989-1993, and 1994-1998. Each ASM panel starts 2 years after the most recent economic census.

[10] The Levinsohn and Petrin (2003) productivity estimation is performed at the two-digit industry level.

investment, shipments, value added, employment, materials use, and wages.[11]

The CMF is collected every 5 years in years ending in 2 and 7, and is the primary source of information about the structure and functioning of the manufacturing sector in the U.S. The manufacturing universe includes around 400,000 establishments. During Censal years, the ASM sample is a subset of the CMF.

As mentioned previously, the sample frame of the ASM is not conducive for the exact timing of the births and deaths of establishments in the universe of manufactures, as not only is it a subsample of the universe of plants, but 80 percent of the establishments are rotated in and out of the sample every 5 years. That said, by design the ASM intends capture births and deaths of establishments within a 5 year sample frame (by tracking the business register), and when weighted by the inverse of their probability weights, entry and exit should in fact be representative of the manufacturing sector as a whole. The remaining problem is to properly label plant entry and exit accordingly during years in which the sample frame rotates. For this purpose we employ the LBD.

The LBD contains the universe of all U.S. business establishments with paid employees from 1976 to present (see (Jarmin and Miranda, 2002)). The LBD contains information on establishment entry and exit, gross job flows, and changes in the structure of the U.S. economy. We link the LBD to our panel of ASM establishments and use the birth death information to properly indicate plant status for the productivity decompositions.

Capital stock data is calculated via the perpetual inventory method and sources from Census microdata and the NBER-CES productivity database. Initial values for buildings and machinery are taken from the Census of Manufacturers buildings at beginning and machinery at beginning variables. Investment for both buildings and machinery are from the ASM data. Investment and starting capital are deflated using price indices from the NBER productivity database.

Table (1) shows the number of plants in our sample. Panel (1a) shows the plant count by year and panel (1b) presents the plant count by two digit industry. The input elasticities estimated using the (Wooldridge, 2009) and (Levinsohn and Petrin, 2003) estimators can be found in table (2).

For the aggregates we forecast, we used data from the BLS Multifactor Productivity (MFP) releases for the manufacturing and nonfarm business sectors.[12] The two aggregates use different approaches to measure multifactor productivity, or total factor productivity.

Industry affiliation changes dramatically over our sample of interest in 1987 (from the 1972 SIC basis). Industry affiliation is concorded to the 1987 Standard Industrial Classification. The 1997 switch to NAICS is one reason our sample ends in 1998.

[11] Nominal values are deflating using the NBER productivity database, which contains shipments, investment, and materials deflators; wages are deflated using the CPI. See (Bartelsman *et al.*, 2000).

[12] For our analysis, we use the superceded historical SIC measures for manufacturing, which contain data from 1949-2001. For the non-manufacturing sector, the BLS MFP historical 1948-2008 Net Multifactor Productivity and Cost tables were used.

Table 1: Plants in sample.

Panel 1a: by year		Panel 1b: by industry		
year	count	industry	SIC Description	count
1974	62,666			
1975	63,643			
1976	63,624	Durable Manufacturing		
1977	67,722	20	Food and kindred products	153,967
1978	65,925	21	Tobacco manufactures	1,702
1979	53,416	22	Textile mill products	48,355
1980	53,049	23	Apparel and other textile products	91,541
1981	52,466	24	Lumber and wood products	94,371
1982	52,816	25	Furniture and fixtures	38,377
1983	48,361	26	Paper and allied products	58,585
1984	53,069	27	Printing and publishing	114,806
1985	52,690	28	Chemicals and allied products	92,620
1986	52,375	29	Petroleum and coal products	20,584
1987	53,034	30	Rubber and miscellaneous plastics products	76,370
1988	50,583	31	Leather and leather products	12,653
1989	54,330			
1990	64,635	Non Durable Manufacturing		
1991	67,019	32	Stone, clay, glass, and concrete products	72,889
1992	66,057	33	Primary metal industries	52,560
1993	62,632	34	Fabricated metal products	149,068
1994	59,091	35	Industrial machinery and equipment	159,732
1995	60,725	36	Electrical and electronic equipment	92,208
1996	65,624	37	Transportation equipment	52,802
1997	63,503	38	Instruments and related products	43,581
1998	54,087	39	Miscellaneous manufacturing industries	36,371

Source: Annual Survey of Manufacturers and Census of Manufactuers, U.S. Census Bureau.

Table 2: Estimated elasticities of labor and capital.

industry	β_l	β_k	se(β_l)	se(β_k)	obs
20	0.410	0.523	0.008	0.012	93,900
21	0.547	0.657	0.078	0.129	1,110
22	0.501	0.324	0.009	0.016	28,880
23	0.555	0.353	0.010	0.011	44,294
24	0.577	0.370	0.008	0.010	55,846
25	0.521	0.360	0.012	0.015	23,243
26	0.481	0.389	0.010	0.014	30,816
27	0.695	0.306	0.009	0.010	67,454
28	0.362	0.507	0.009	0.014	59,030
29	0.335	0.480	0.014	0.035	12,602
30	0.424	0.461	0.008	0.016	31,970
31	0.565	0.332	0.023	0.027	7,786
32	0.535	0.421	0.006	0.012	45,384
33	0.531	0.435	0.009	0.017	31,188
34	0.551	0.368	0.005	0.009	92,461
35	0.592	0.376	0.006	0.009	93,897
36	0.389	0.506	0.008	0.016	50,906
37	0.644	0.351	0.011	0.015	33,384
38	0.438	0.485	0.012	0.018	24,393
39	0.541	0.357	0.011	0.015	19,845

The production function parameter estimates are derived as in (Wooldridge, 2009) and using the number of employees for the labor input. Output is measured by value added. Source: Census of Manufactures, Annual Survey of Manufactuers, and NBER Manufacturing Productivity Database.

For the manufacturing sector, the release contains real output and inputs on a KLEMS-basis for capital, labor, energy, materials, and business services. From this data source, we use the value added, sectoral output, capital, labor and multifactor productivity series from

1974 to 1998. The BLS calculates MFP as the ratio of the sectoral output index to an index of the combined inputs of labor, capital services, energy, non-energy materials, and purchased business services. The BLS's framework for the manufacturing sector uses sectoral output as the output measure, i.e., the value of production less that portion which is consumed in the same industry. The input composite is calculated as a weighted average of individual inputs, where the weights are the shares of each input in current dollar output.

For private nonfarm business,[13] the BLS data contains real output, and labor and capital inputs from 1974 to 1998. The multifactor productivity indexes are derived by dividing an output index by an index of labor input and capital services. The output measure for nonfarm business is value added based, i.e., the exclusion of intermediate transactions between businesses.

Aside from different input and output measures used to estimate MFP by the BLS for the manufacturing and nonfarm business sectors, the input measures are calculated similarly for both MFP frameworks. Labor is measured as the hours worked by all persons engaged in a sector. The sources for employment and average weekly hours data are the BLS Current Employment Statistics program and Current Population Survey. Capital input is defined as the flow of services from physical assets, which include equipment, structures, inventories and land.[14] We also utilize labor quality, or labor composition, the ratio of labor input to hours of all persons.[15]

2.2 Aggregates

We now turn to the aggregates of the microdata used in the analysis.[16] Before analyzing the efficacy of forecasting using micro-aggregates, it is important to compare how the aggregated microdata compare to published aggregates. Figure 1 contains comparison plots of value added growth with the growth in BLS output measures and a comparison of aggregates of total factor productivity with the BLS multifactor productivity estimates.

First, figure (1a) contains a comparison of value added from aggregated Census microdata used in our sample with the published BLS aggregate. The two series track each other fairly closely and have a correlation coefficient of 0.9. Unfortunately, while value added is the measure we utilize to estimate plant-level TFP, value added is not the output measure used by the BLS when calculating official multifactor productivity estimates for the manufacturing sector. The BLS employs sectoral output in their KLEMS approach for the U.S. manufacturing sector.[17]

In response, we also plot and correlate our aggregated value added measure to the BLS

[13] The private nonfarm business sector is GDP less government, less the output of household workers, nonprofits, and the farm sector.

[14] For more information see (Gullickson and Harper, 1987).

[15] Labor composition measures the effect of shifts in age, education, and gender composition on the work force.

[16] The productivity estimation techniques are described in the following section.

[17] The KLEMS approach using sectoral output allows for the use of Domar weights to aggregate up subindustries into total factor productivity for the entire economy or a particular sector.

sectoral output measure in figure (1b) and find a similarly high correlation of 0.88. Our aggregation of Census microdata captures the dynamics of the manufacturing sector relative to the published aggregates well. While both metrics of output are derived from the same source data, it is not *fait accompli* that the published aggregate should closely track the series constructed from Census microdata.

This conclusion is particularly true when thinking in terms of comparing our aggregates of manufacturing total factor productivity with the published total factor productivity estimates from the BLS. This is because the two methodologies differ in their approach. We use the Wooldridge-Levinsohn-Petrin approach to estimate TFP and aggregate up using value added or input weights. Alternatively, the BLS estimates of total factor productivity growth are defined as the difference between output growth and the growth of a composite of the aggregate level of inputs. In this case a weighted combination of capital, labor, energy, materials, and business services. The BLS and our approaches to net out input contributions differ substantially. Nonetheless, as seen in figure (1c), the micro-aggregated TFP series tracks the BLS published estimates fairly closely, with a correlation of 0.81.

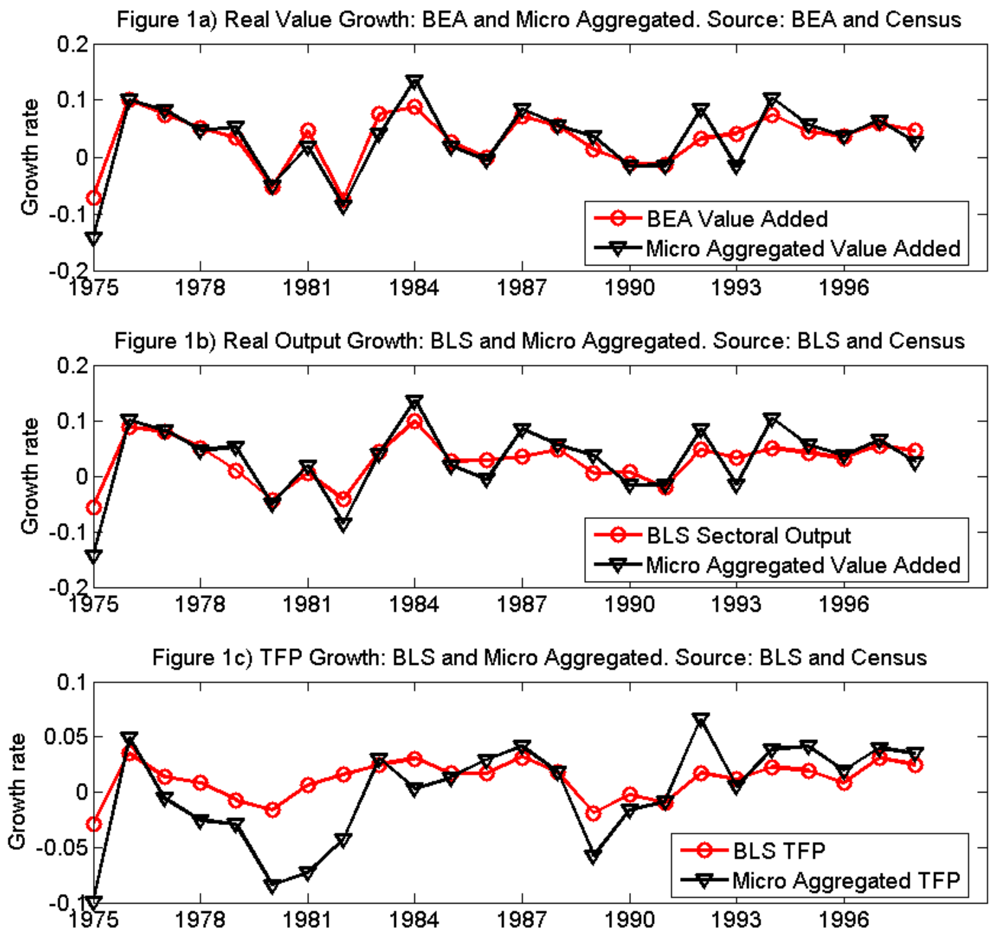

Figure 1: Plots of Microaggregated Data and Published Aggregates.

3 Firm-level TFP measures and productivity decompositions

We now describe the microeconometric work conducted in a secure computing environment at the Census Bureau using confidential plant-level information. The output of this work, to be used in the forecasting exercises, is a collection of micro-aggregated productivity components, or firm-level information on productivity and size aggregated into time series components following productivity decomposition accounting rules. We start by briefly describing the methods used to estimate firm-level TFP measures, and then show the productivity decompositions used to aggregate firm-level information into micro-aggregated productivity components.

The production function for firm i at time t we employ is Cobb-Douglas of the form:

$$y_{it} = \beta_0 + \beta_l l_{it} + \beta_k k_{it} + \tau_{it} + \epsilon_{it},$$

where y_{it} is log value-added, k_{it} is log capital, l_{it} is log labor, τ_{it} is log productivity and is ϵ_{it} assumed to be an *iid* disturbance. τ_{it} is unobservable by the econometrician but known to the firm. Since τ_{it} is in the information set on which the firm conditions its optimal choices of inputs, there will always be a non-negative correlation between input factors and τ_{it}. This dependence biases simple OLS parameter estimates. To estimate TFP at the firm-level we apply a standard semi-parametric procedure introduced by (Olley and Pakes, 1996) and modified by (Levinsohn and Petrin, 2003) and (Wooldridge, 2009) (OP and LP henceforth).

We estimate total factor productivity for both plants and firms using the Wooldridge-Levinsohn-Petrin estimator to control for simultaneity between unobservable productivity and the observable input choices. This approach solves the simultaneity problem by using a proxy for the transmitted plant-specific efficiency. The specific intermediate input chosen as a proxy for the unobservable shock is energy. The Wooldridge approach to estimating LP productivity implements the moment conditions in a Generalized Methods of Moments framework, corrects for the simultaneous determination of inputs and technical efficiency, and is robust to the (Ackerberg *et al.*, 2005) critique.

3.1 Productivity Decompositions

The micro-aggregrated components we employ are decompositions based on accounting identities. We start by looking at the decomposition of (Baily *et al.*, 2001) and continue by building on a decomposition given by (Olley and Pakes, 1996). The two different decompositions are aggregated using four different weighting schemes allowing for eight different sets of micro-aggregated components. First, the components may be computed using the two accounting identities (or productivity decompositions) described below by equations (1) and (5). Next, the firm-level 'market share' measure that is used in the decomposition may be based on the firm's share of industry value added, or the firm's share of industry inputs. Finally, the micro-aggregated components are generated using either fitted or actual values of firm-level market

share and firm-level productivity. The fitted values are estimated using auxiliary econometric models to capture productivity 'push and pull' effects.[18]

3.1.1 Accounting Identities

The first dynamic decomposition is derived by taking a total derivative of the share-weighted change in total factor productivity and is based on (Griliches and Regev, 1995) as formulated in (Baily *et al.*, 2001). The change in total factor productivity is decomposed as:

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

where τ_t is the aggregate TFP-level, τ_{it} is the TFP-level of firm i in period t, ϕ_{it} is the market share of firm i in period t, $\widetilde{\tau}$ is the moving average of aggregate productivity in time t given by $\widetilde{\tau} = \frac{\tau_{t-1}+\tau_t}{2}$, Δ is the difference operator, C denotes the set of continuers, E denotes the set of entrants and X denotes the set of exiters in time t. The four terms in the previous equation are called within-, between-, entry-, and exit-terms.

In the second decomposition, a static decomposition is combined with equation (1). The static equation was introduced by (Olley and Pakes, 1996). The static decomposition is

$$\tau_t = \overline{\tau}_t + \sum_{N_t}(\phi_{it} - \overline{\phi}_t)(\tau_{it} - \overline{\tau}_t), \tag{2}$$

where $\overline{\tau}_t = \frac{1}{N_t}\sum_{N_t}\tau_{it}$, and $\sum_{N_t}(\phi_{it} - \overline{\phi}_t)(\tau_{it} - \overline{\tau}_t)$ is the covariance between productivity and market share. This decomposition informs about the joint distribution of firm-level productivity and firm size. When the covariance-term is high, aggregate productivity is higher than the unweighted average of firm-level productivity.

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 (2) yields:

$$\Delta\tau_t = \Delta\overline{\tau}_t + \sum_{N_t}(\phi_{it} - \overline{\phi}_t)(\tau_{it} - \overline{\tau}_t) - \sum_{N_{t-1}}(\phi_{it-1} - \overline{\phi}_{t-1})(\tau_{it-1} - \overline{\tau}_{t-1}), \tag{3}$$

i.e. aggregate productivity change is the sum of the change in the unweighted average productivity and the change in the covariance term. Entry and exit are implicitly accounted for in equation (3) as the sums run on $i = 1...N_t$ and $i = 1...N_{t-1}$. Hence, the weights are such that

[18] These simple models generate fitted values for individual firms' market shares and productivity growth, which, if the models have strong explanatory power, may add to aggregate forecast performance. See appendix A.1 for further information on the estimation of the fitted values.

$\sum_{N_t} \phi_{it} = 1$ and $\sum_{N_{t-1}} \phi_{it-1}$. Rewrite $\widetilde{\tau}$ using (2):

$$\begin{aligned}\widetilde{\tau} &= \frac{\tau_t + \tau_{t-1}}{2} = \frac{\overline{\tau}_t + \overline{\tau}_{t-1}}{2} + \frac{cov_t(\phi,\tau) + cov_{t-1}(\phi,\tau)}{2} \\ &= \widetilde{[\overline{\tau}_t]} + \widetilde{cov}(\phi,\tau). \end{aligned} \tag{4}$$

We base our analysis on (1), but we also want to introduce information about the size and productivity distributions. We expect $\widetilde{cov}(\phi,\tau)$ to do this. Plug (4) into the between term of (1) to get

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

so (1) looks like

$$\begin{aligned}\Delta\tau_t &= \\ &\sum_C \widetilde{\phi}_i\Delta\tau_{it} + \sum_C \Delta\phi_i(\widetilde{\tau}_i - \widetilde{[\overline{\tau}_t]}) - \sum_C \Delta\phi_i\widetilde{cov}(\phi,\tau) \\ &+ \sum_E \phi_{it}(\tau_{it} - \widetilde{\tau}) - \sum_X \phi_{it-1}(\tau_{it-1} - \widetilde{\tau}).\end{aligned} \tag{5}$$

The between-term component of (1) now has two terms: the first sums share-changes weighted by deviations from the time-average of the simple cross-section average and the second shows the effect of the covariance-term.[19]

4 Univariate and Multivariate Forecasting

We now estimate and forecast productivity. We first employ univariate and multivariate forecasts, and then turn to a state space framework. The univariate and multivariate forecasting techniques follow the methodology found in (Bartelsman and Wolf, 2009). We evaluate our forecasts in four three-year rolling forecast windows.[20] In each horse-race, we compare forecast performance using only the aggregate timeseries with the forecast performance of a set of micro-aggregated components. The benchmark for our comparisons is the average growth of the Hodrick-Prescott-trend of aggregate TFP for a particular 3-year window. When comparing two forecasts we calculated the following performance metric:

$$\widehat{\triangle\tau}^{Mi}_{t+s} - \overline{\triangle\tau}^{HP}_{t+s} \text{ vs } \widehat{\triangle\tau}^{Ag}_{t+s} - \overline{\triangle\tau}^{HP}_{t+s}, \tag{6}$$

[19] The ϕ_{it} weights were defined at the beginning such that $\sum_{N_t} \phi_{it} = 1$, hence $\sum_C \Delta\phi_{it}$ is not necessarily zero. It would be zero had we defined $\sum_C \phi_{it} = 1$. Therefore, the covariance term is not cancelled out by $\sum_C \Delta\phi_i$.

[20] We evaluated our models in the following forecast windows: 1993-1995, 1994-1996, 1994-1997, 1995-1998.

where $\widehat{\triangle\tau}^{Mi}_{t+s}$ denotes the forecast using micro-aggregated components, $\widehat{\triangle\tau}^{Ag}_{t+s}$ denotes the forecast of the aggregate timeseries, $\overline{\triangle\tau}^{HP}_{t+s}$ denotes the average growth of the Hodrick-Prescott-trend of the aggregate, and s denotes the forecast window in which the forecast is evaluated.

4.1 Univariate forecasts

To start we employ a straightforward univariate framework. For univariate forecasts, the definition of $\widehat{\triangle\tau}^{Mi}_{t+s}$ is implied by the productivity decompositions found in section 3.1, which parse aggregate productivity growth into micro-aggregated components:

$$\widehat{\triangle\tau}^{Mi}_{t+s} = \sum_{i\in\{w,b,ne\}} \widehat{\triangle\tau}^{i}_{t+s},$$

where $\{w, b, ne\}$ denotes the set of within, between and net entry components. $\widehat{\triangle\tau}^{i}_{t+s}$ denotes the forecast from separate univariate autoregressive $AR(p)$ specifications for component i.[21] $\widehat{\triangle\tau}^{Ag}_{t+s}$ is also calculated from a univariate autoregressive specification. Finally, we compare how far off the estimates of $\widehat{\triangle\tau}^{Ag}_{t+s}$ and $\widehat{\triangle\tau}^{Mi}_{t+s}$ are from $\overline{\triangle\tau}^{HP}_{t+s}$, as in (6).

4.2 Multivariate Forecasts

We utilize a more complex framework for our multivariate forecasts. The framework entails forecasting actual published productivity aggregates and pulls in information from other related macroeconomic data. For the multivariate forecasting approach, we calculated $\widehat{\triangle\tau}^{Mi}_{t+s}$ and $\widehat{\triangle\tau}^{Ag}_{t+s}$ using vector-autoregressions (VARs). We define the vector of endogenous variables in the VAR as $y_t = (\triangle va_t,\, \triangle k_t,\, \triangle l_t,\, \triangle\tau^{Ag}_t)$, i.e. a vector of logarithmic differences of value added, capital, labor and aggregate TFP. We denote the regressand by $y_t = (\triangle va_t,\, \triangle k_t,\, \triangle l_t,\, \triangle\tau^{Ag}_t)$.[22] The regressors include some of y_t's own lags and also lagged micro-aggregated components. We denote the latter as $\triangle\tau^i_{t-q}$. Our data for y_t source from the BLS, while the $\triangle\tau^i_{t-q}$ were constructed from firm-level data. In short, and using the generic notation $VAR(p, q)$ for a VAR with p endogenous lags of y_t and q lags of predetermined micro-aggregated components $\triangle\tau^i_t$, we estimated VARs with the following pairs of p and q: $(1, 2), (2, 1), (1, 1)$. The corresponding aggregate VARs do not feature $\triangle\tau^i_{t-q}$ as additional regressors. Hereafter, we refer to the first set of VARs as microcomponent-VARs and to the latter as aggregate-VARs.

We outline the Bayesian forecasting approach assuming we have retrieved a posterior distribution for the coefficients and error covariance matrix of a VAR. Denote the posterior estimates

[21] The lag order of the $AR(p)$ $p \in \{1, 2\}$ models based on the BIC, which selects p=1. The lag order was restriced to one or two lags in order to preserve degress of freedom.

[22] The components of y_t are implied by the production function.

by β and V, respectively. The main object of interest is the predictive density

$$p(y_{T+1}|y_T, M) = \int_0^\infty \int_{-\infty}^\infty p(y_{T+1}|V, \beta, M)p(V, \beta|y_T, M)d\beta dV, \tag{7}$$

where y_T is observed data up to time s, y_{T+s} is the forecast s-period ahead, V is the covariance matrix of shocks and β is the parameter matrix in the VAR, and M denotes the forecast model. The predictive density integrates (i) the uncertainty about β, V and (ii) the intrinsic uncertainty about the future value of y_{T+1}, conditional on the history y_T of observed data and model M. We retrieved the joint posterior $p(V, \beta|y_T, M)$ and the predictive density $p(y_{T+s}|V, \beta, M)$ from a customized Gibbs-sampler.[23]

We choose a Bayesian approach for our multivariate forecasting exercise, as it allows us to compare the fit of different specifications even if they are not nested. This is an important advantage because model selection is an important issue when micro-aggregated components are combined with macro-information in a multivariate framework. Furthermore, averaging forecasts in this framework is also relatively straightforward, which is important for our purposes as model selection proved to be critical. We computed the Bayesian Model Average (BMA) of forecasts over the entire set of VAR specifications in a forecast window. The average was taken over twenty-four specifications[24] and the weights were based on each specification's Predictive Bayes Factor (PBF). Deviations from average HP-trend growth were computed using the means of the forecast-distributions of individual growth trajectories over the forecast period, not the means of the yearly distributions of forecasts.

4.3 Results

The univariate forecasting results can be found in the first two columns of table (3). The first column presents the forecasts based on micro-aggregated components and the second column presents forecasts based on the aggregate timeseries. Aside from the third forecasting window all forecasts using micro-aggregated components from the productivity decompositions outperform aggregate forecasts. For example, the column labelled "Microcomponents" for the 1992 to 1994 forecast window shows that the aggregate of the forecasts of the micro-aggregate components is 1.5 percentage point lower than the average of the HP-trend. The forecast of aggregate TFP is 1.9 percentage points lower than the HP-trend for the period, and thus performs 0.4 percentage point worse. Our results imply that forecasts including micro-aggregated components detect more of the productivity acceleration in the nineties, relative to the aggregate forecasts.

The last two columns of table (3) summarize the result of Bayesian model-averaging when the forecast target is the published aggregate TFP. The microcomponent-VAR forecasts dom-

[23]See, e.g. (Koop, 2003).

[24]The dimensions are: i) two decompositions ii) input-factor or value added shares and their fitted values (four-scenarios) iii) three lag-specifications. These yielded 24 microcomponent-VARs and three aggregate-VARs in each forecast window.

Table 3: Out-of-sample forecast performance in different forecast periods, percentage point deviation from HP-trend of aggregate manufacturing productivity growth. Wooldridge estimates, using Employment.

	Univariate Forecast		BMA-VAR Forecast	
	Microcomponents	Aggregate	Microcomponents	Aggregate
1992-1994	-1.5	-1.9	-0.9	-1.1
1993-1995	-1.4	-1.9	-0.6	-0.9
1994-1996	-1.3	-1.0	-0.9	-1.4
1995-1997	-1.4	-1.7	-0.6	-1.2

Note: Results for the univariate forecast were calculated using input factor shares of firms. These factor-shares are based on the weighted average of input factors. The columns should be read as a deviation from Hodrick-Prescott trend over the time frame of interest. Negative values are read as falling below trend. Weighted average forecasts were calculated using 24 VAR models for microcomponent VARs and three VAR models for the aggregate VARs. The weights are based on the Predictive Bayes Factor of forecast models. The specification set of the microcomponent VARs is spanned by three dimensions: 1) two decompositions 2) input-factor or value added shares and their fitted values (four-scenarios) 3) three lag-specifications for each VAR(p,q): ((1,2),(2,1),(1,1)). For instance, a VAR(1,2) includes the 1st lag of endogenous variables (growth rates of aggregate capital services, labor services, tfp, value added), the 1st and 2nd lags of predetermined variables. Predetermined variables are lagged values of micro-aggregated components. The specification set for aggregate VARs consists of three VARs without microcomponents in each forecast window. The three VARs are given by the three lag-specifications above.

inate the aggregate VAR forecasts in every forecast window. Of note, in the last two forecast windows the micro-aggregated component forecasts are about $\frac{1}{2}$ percentage point higher than the aggregate VAR forecasts. In that same period our estimates of average trend growth in total factor productivity was 1.8 percent. As a result, the aggregate VAR forecasts would have missed roughly a third of the late-nineties productivity acceleration. We now turn to using a more structural framework to estimate total factor productivity.

5 Using micro-aggregated components to estimate structural productivity

The addition of micro-aggregated components to both univariate and multivariate forecasting experiments improves our ability to forecast and understand trend productivity in the manufacturing sector. A natural extension would be to see if the within, between and net entry terms aid in forecasting aggregate productivity growth for the economy as a whole. This is important, particularly in macroeconomic forecasting, as estimates of potential output and the application of Okun's law depend on structural, or trend, productivity.

Unfortunately, a longitudinally linked plant-level dataset with sufficient information in order to decompose productivity for the entire economy over a long enough time horizon does not exist. Given the data constraints, this leaves the manufacturing sector as our sole source of micro-aggregated components. *Prima facie*, the manufacturing sector seems like an inadequate starting point for inferences about understanding cyclical dynamics of the overall economy, particularly given that the sector represents roughly only 10 percent of total U.S. output. That said, despite the relative size of the sector, output growth in the manufacturing sector has

actually kept pace with the rest of the economy,[25] while exhibiting greater cyclical swings.

As a result, the sector tends to make outsized contributions to GDP growth during economic turning points.[26] Moreover, the manufacturing sector contributes an outsized portion of overall business cycle volatility. In particular, (Ramey and Vine, 2006) find that roughly a quarter of the variance in GDP sources from the motor vehicle sector, even though the industry accounts for less than 4 percent of the level of GDP. In this environment, using the micro-aggregated components from the manufacturing sector to extract a signal about the overall trend in aggregate productivity might actually prove useful. This conjecture proves somewhat correct. The following section provides evidence that the addition of micro-aggregated components provides insight into structural multifactor productivity.

5.1 Structural productivity and potential output

Currently, most forecasts of trend TFP and potential output continue to be based on the representative firm view and thus only need to worry about measurement at the aggregate level. These practices do not consider firm-level heterogeneity, technology diffusion, shifts in market share, and appropriate aggregation methods. TFP is estimated typically within a production function equation, which is used to estimate potential output. Concrete applications differ as to how potential output and TFP is modeled. For instance, one of the Federal Reserve's models for potential output defines TFP as unobserved stochastic process using a state space framework.[27] At the European Central Bank, only in-sample trend TFP values are calculated by applying the Hodrick-Prescott filter to the residual of the production function, TFP is not extrapolated directly.[28]

Improving the estimates of productivity trends should ultimately lead to better macroeconomic forecasts through improved measures of potential output, i.e., the level of output that would be achieved were factor inputs fully utilized and multifactor productivity at its trend level.

We build on previous methodology to estimate trend multifactor productivity. The following primarily uses the framework of (Roberts, 2001) and (Kuttner, 1994). First, decompose log per capital output q_t into an hours h_t and output per hour π_t:

$$q_t = h_t + \pi_t.$$

[25] From 1960 to 2009, the average annual rate of change in real nonfarm business output was 3.5 percent, only slightly higher than the 3.2 percent annual change for manufacturing. The relatively faster gains in manufacturing productivity have resulted in lower relative goods prices which, in combination with inelastic demand for goods (on average), has led to a decline in manufacturing's share of nominal output.

[26] See (Corrado and Mattey, 1997) for details.

[27] See (Kuttner, 1994) for a method of estimating potential output as an unobserved trend. Other benefits of the state space approach include updating the trend in real-time, in response to data releases and the ability to assign confidence intervals to estimates of trend.

[28] See (Fagan *et al.*, 2001) for more details.

In a trend-cycle model, variable x_t is assumed separable in its trend and gap components x_t^* and $xgap_t$. We adopt the separability assumption and assume that $q_t = q_t^* + qgap_t$, $h_t = h_t^* + hgap_t$ and $\pi_t = \pi_t^* + \pi gap_t$, where the trend variables are unobserved.[29] Our primary focus is on π_t. In order to estimate trend multifactor productivity, the next section describes a decomposition of π_t^* based on a Cobb-Douglas production function.

5.2 Labor productivity and trend total factor productivity

The standard Cobb-Douglas production function is a natural starting point for any model with total factor productivity. Writing output as

$$Q_t = A_t K_t^{\alpha} (E_t H_t)^{1-\alpha},$$

where K_t denotes capital services, E_t and H_t denote labor quality and hours, respectively, and A_t denotes total factor productivity. The production function implies a definition for log-labor-productivity, denoted by π_t and measured as the log of output per hour:

$$\pi_t = a_t + \alpha(k_t - h_t) + (1-\alpha)e_t, \tag{8}$$

where $(k_t - h_t)$ denotes capital deepening. Decomposing labor productivity into a trend and cycle component implies

$$\pi_t = \pi_t^* + \pi gap_t. \tag{9}$$

Given data on both capital deepening and labor quality trends, we can parse π_t^* into trend multifactor productivity (a_t^*), cyclical labor productivity and the two observables, capital deepening $(k_t - h_t^*)$ and labor quality $((1-\alpha)e_t)$ terms. Combining equations (8) and (9) implies

$$\begin{aligned} \pi_t &= \pi_t^* + \pi gap_t \\ &= a_t^* + \alpha(k_t - h_t^*) + (1-\alpha)e_t + \pi gap_t. \end{aligned} \tag{10}$$

Equation (10) serves as the basis for our analysis. Trend multifactor productivity is assumed to be an unobserved component of our state space framework. We estimate trend multifactor productivity using data on manufacturing and the private nonfarm business sector. Our state space model builds on the assumptions in (Roberts, 2001) about the relationship between πgap_t and $hgap_t$. We briefly describe these assumptions in the next section.

5.2.1 A model for hours

In (Roberts, 2001), the cyclical component of current hours is assumed to depend its own lagged values and on the cyclical component of output. Implicit in this assumption is a partial

[29] These definitions imply that we can write potential output as $q_t^* = h_t^* + \pi_t^*$.

adjustment model where a cyclical productivity-shock yields a less-than-proportional response in hours on impact. In the wake of a shock, productivity increases above trend, then hours gradually adjust until the productivity gap returns to zero. The catch-up idea is encapsulated by including lagged values of πgap_t and $qgap_t$ in the equation for $hgap_t$:

$$hgap_t = \theta_0 qgap_t + \theta_1 qgap_{t-1} + \theta_2 hgap_{t-1} + u_{2t}. \tag{11}$$

The idea of partial adjustment can be represented by

$$\begin{aligned} \Delta hgap_t &= \mu_0 \Delta qgap_t + \mu_1 \pi gap_{t-1} + u_{2t} \\ 0 &< \mu_0 < 1, \mu_1 > 0, \end{aligned} \tag{12}$$

which is equation (12) in Roberts (2001). It can be shown that (12) is a special case of (11) with restrictions

$$\theta_0 = \mu_0, \theta_1 = \mu_1 - \mu_0, \theta_2 = 1 - \mu_1.$$

5.2.2 Discussion

Firm-level information may help explaining aggregate productivity trends. An obvious test of this idea in the current framework would be to see what difference micro-aggregated components make when used in the estimation of the level of trend TPF and its growth rate. A full-information maximum likelihood approach would ideally involve specifying the data generating processes both for the trend-component of hours and the cyclical component for labor productivity. However, we work with a simpler approach and assume that we have estimates[30] for trend hours (h_t^*). We treat πgap_t as an unobserved variable and estimate its value.

We deviate from (Roberts, 2001) for two reasons. First, our focus is on whether micro-aggregated components help explain changes in a_t^*, and it is not clear how Roberts (2001) framework accommodates an equation describing the relationship between the micro-aggregated components and the change in a_t^*. Second, we want to facilitate parameter estimation by applying the Expecation-Maximization algorithm (EM henceforth), which combines analytical expressions and numerical procedures.

As we will see, the state space specification presented in the next section directly accommodates our micro-aggregated components. It also allows us to maximize the complete data likelihood. Maximizing the complete data likelihood amounts to considering the joint density of the observation and state vectors simultaneously, and it is therefore closer to a full-information approach. The EM finds the maximum point using both numerical optimization and using analytical results. The analytical expressions facilitate numerical search and therefore convergence is achieved faster.

[30]We use the Hodrick-Prescott filter with $\lambda = 100$.

As final remark, we generate micro-aggregated components using 25 years of firm-level data at an annual frequency. The number of observations may appear relatively small at first glance. We acknowledge that the majority of state-space applications use longer timeseries or data at higher frequencies.[31] However, it is not a priori obvious whether the time-span of our sample or the yearly frequency prevents us from precisely estimating productivity trends.

5.3 A state-space representation

Define the labor productivity residual as log-labor-productivity (π_t) minus capital deepening $\alpha(k_t - h_t^*)$ and labor quality $(1-\alpha)e_t$:

$$\pi_t^r = \pi_t - \alpha(k_t - h_t^*) - (1-\alpha)e_t. \tag{13}$$

Equation (10) shows that the labor productivity residual can be decomposed into trend term and cyclical productivity-gap terms:

$$\pi_t^r = a_t^* + \pi gap_t, \tag{14}$$

On the other hand, the productivity-gap is also related to the hours-gap. First, substitute out π_t using equation (8):

$$\begin{aligned}
\pi_t^r &= a_t + \alpha(k_t - h_t) + (1-\alpha)e_t \\
&\quad - \alpha(k_t - h_t^*) - (1-\alpha)e_t \\
&= a_t - \alpha(h_t - h_t^*) \\
&= a_t - \alpha * hgap_t.
\end{aligned} \tag{15}$$

Combining equations (14) and (15), we have the following relationship between πgap_t and $hgap_t$:

$$\pi gap_t = a_t - a_t^* - \alpha * hgap_t. \tag{16}$$

Equation (16) says that the cyclical component of labor productivity is a linear function of trend TFP and the hours-gap. In this sense, our specification is similar to the approach in (Roberts, 2001) (see equation (12)). (Roberts, 2001) estimates trend productivity on quarterly data. At this frequency, it is conceptually easy to justify the catch-up idea and relate movements in hours to lagged productivity levels. At yearly frequency, we observe these movements simultaneously, which renders identification more difficult. However, we can *assume* that πgap_t is predetermined with respect to $hgap_t$, as suggested by partial adjustment. Our assumption implies that if we observe an increasing hours gap in a year, we can infer that the productivity gap must have

[31] For example, (Roberts, 2001) employs quarterly timeseries starting in the early 1950's to estimate total factor productivity trends.

decreased. Equation (16) allows us to use $hgap_t$ to explain the variation in πgap_t. We also allow πgap_t to depend on πgap_{t-1}.

We do not impose further structure on the relationship between $hgap_t$ and πgap_t. Any restriction implied by the gradual adjustment idea would yield rank reduction in the state space representation. We want to avoid this for computational reasons, because we want to develop our estimation routine based on earlier results in the literature.[32]

We denote micro-aggregated components by c_{it}, $i = 1, 2, 3$, and assume that they signal about trend productivity growth. We denote trend productivity growth by γ_t, and adopt the assumption in (Roberts, 2001) that γ_t follows a random walk

$$\begin{aligned}\gamma_t \equiv \Delta a_t^* &= a_t^* - a_{t-1}^* \\ &= \gamma_{t-1} + e_{\gamma_t}.\end{aligned}$$

5.3.1 A model in first differences

We estimate the model in first-differences for three reasons. First, the level series exhibit a strong trends, which is taken up by the filtered estimate of a_t^*, possibly leading to less accurate estimates of the growth rate of the trend γ_t. Second, the micro-aggregated components are by definition contributions to the growth rate of aggregate productivity, therefore we expect them to inform about trend growth rather than the level. Finally, there are fewer parameters in the first-differenced model, which leaves more degrees of freedom for estimation.

We obtain the observation equation by differencing the labor productivity residual in equation (13):

$$\Delta\pi_t^r = \Delta\pi_t - \alpha\Delta(k_t - h_t^*) + (1-\alpha)\Delta e_t.$$

Differencing a_t^* yields $\gamma_t + e_{\gamma_t}$ implying that $\Delta\pi_t^r = \gamma_t + \Delta\pi gap_t + u_t$. Casting these equations into state space form gives

$$\Delta\pi_t^r = (1\ 1)\begin{bmatrix}\gamma_t \\ \Delta\pi gap_t\end{bmatrix} + v_t, \tag{17}$$

and

$$\begin{aligned}\begin{bmatrix}\gamma_t \\ \Delta\pi gap_t\end{bmatrix} &= \begin{bmatrix}1\,0 \\ 0\,\rho\end{bmatrix}\begin{bmatrix}\gamma_{t-1} \\ \Delta\pi gap_{t-1}\end{bmatrix} \\ &\quad + \begin{bmatrix}b_1\,b_2\,b_3\ 0 \\ 0\ \ 0\ \ 0\ b_4\end{bmatrix}\begin{bmatrix}c_{1t-1} \\ c_{2t-1} \\ c_{2t-1} \\ \Delta hgap_t\end{bmatrix} + \begin{bmatrix}e_{\gamma_t} \\ e_{\Delta\pi gap_t}\end{bmatrix}.\end{aligned} \tag{18}$$

[32] See section 5.5 for more details on this issue.

Now consider the general linear model

$$\begin{aligned} y_t &= C * x_t + v_t \\ x_t &= A * x_{t-1} + B * u_{t-1} + w_t \\ v_t &\sim NID(0, R) \\ w_t &\sim NID(0, Q) \\ x(0) &\sim N(init\_x, init\_V), \end{aligned} \tag{19}$$

where $y_t \in \mathbb{R}^p$ denotes the observation vector, $x_t \in \mathbb{R}^k$ denotes the unobserved state vector. Variable u_t is referred to as the control signal. $A \in \mathbb{R}^{k\times k}$, $B \in \mathbb{R}^{k\times s}$ and $C \in \mathbb{R}^{p\times k}$ denote the system matrices and $x(0)$ denotes the initial value of the state vector. In our case, $p = 1$, $k = 2$ and $s = 4$.

It is immediate that our representation in equations (17)-(18) correspond to the general state-space model in (19). The micro-aggregated components enter the system in u_{t-1}. Their effect on trend productivity growth is estimated by coefficients b_1, b_2 and b_3. The equations of the first-differenced model can be cast into the state-space form in (19) with the following definitions

$$\begin{aligned} & A \equiv \begin{bmatrix} 1\,0 \\ 0\,\rho \end{bmatrix}, B \equiv \begin{bmatrix} b_1\, b_2\, b_3\; 0 \\ 0\;\; 0\;\; 0\; b_4 \end{bmatrix}, C \equiv (1\ 1) \\ & x_t \equiv (\gamma_t, d\pi gap_t)', u_t \equiv \left(c_t^w, c_t^b, c_t^{ne}, dhgap_t\right)', w_t \equiv (e_{\gamma_t}, e_{d\pi gap_t})' \\ & y_t \equiv \pi_t^r = \pi_t - \alpha(k_t - h_t^*) - (1-\alpha)e_t \\ & w_t \sim NID(0, Q),\ Q = \begin{bmatrix} \sigma_\gamma^2 & 0 \\ 0 & \sigma_{d\pi gap}^2 \end{bmatrix} \\ & v_t \sim NID(0, R),\ R = \sigma_v^2. \end{aligned} \tag{20}$$

5.4 Parameter constraints and model selection

It is not without purpose that we *defined* the system matrices above. Choosing them as in (20) allows us to represent the equations of the underlying structural model in the general state-space form in (19). In other words, the parameter constraints within those matrices are also implied by the structural model. Therefore, we treat those constraints as identifying restrictions.[33]

One implication of this approach is that we do not formally test their validity. In principle, the likelihood-ratio test would be a standard tool to compare and choose between a model with constraints and one without constraints. However, if we calculate likelihood using identifying restrictions, then we assume that a particular likelihood value can only be calculated under a

[33] Similarly to the identifying restrictions of the SVAR literature.

particular set of constraints. If the constraints are different, the underlying structural model is different and so is the likelihood. This logic implies that we cannot formally test whether one model provides a strictly better or worse fit than another. At best, likelihood values are viewed as a rough-and-ready indicator of how likely that a specific model generated the data.

Unfortunately, there is no panacea to the problem of model selection. We do not want to apply Bayesian or other techniques to compute model averages because we would like to keep the analysis as close as possible to the current macroeconomic practice in estimating structural productivity. We judge a model's fit on how well γ_t, the estimated trend growth rate of TFP, follows known dynamic patterns over the estimation period.[34] We also consider the dynamics of the productivity-gap and "reject" a model if it produces productivity-gap estimates that contradict earlier findings.[35] We are going to discuss these issues in detail in section 5.6. The next section outlines the estimation procedure.

5.5 Estimation

We now describe the basic equations of the EM-algorithm used to estimate the system matrices of the Kalman filter. A detailed description of the algorithm for the general linear model in equation (19) without control signal can be found in (Ghahramani and Hinton, 1996).

The joint density of the state and the observation vectors is given by

$$P(\mathbf{x}, \mathbf{y}) = P(x_0) \prod_{t=2}^{T} P(x_t|x_{t-1}) \prod_{t=1}^{T} P(y_t|x_t), \tag{21}$$

where $\mathbf{x} = (x_1 ... x_T)'$ and $\mathbf{y} = (y_1 ... y_T)'$. The Gaussian model in equation (19) implies that we can write the density of observation y_t, conditional on the state x_t, as

$$P(y_t|x_t) = \exp\left\{ -\frac{1}{2} [y_t - Cx_t]' R^{-1} [y_t - Cx_t] \right\} (2\pi)^{-p/2} |R|^{-1/2}. \tag{22}$$

$P(x_t|x_{t-1})$, the density of x_t conditional on its previous value x_{t-1} is given by

$$\exp\left\{ -\frac{1}{2} [x_t - Ax_{t-1} - Bu_{t-1}]' Q^{-1} [x_t - Ax_{t-1} - Bu_{t-1}] \right\} (2\pi)^{-k/2} |Q|^{-1/2}. \tag{23}$$

Equation (23) contains the control signal, denoted by the u_{t-1} term.

The EM-algorithm amounts to taking logs in (21) and computing the expectation with respect to the data vector $\mathbf{y}$ and differentiating the expectation with respect to the parameters.[36]

[34] In our judgement we also give weight to visual inspection of the dynamics of filtered state variables.

[35] One key property is stationarity, which is implied by the catch-up model. The other one is the ability of signalling the phases of the business cycle in the U.S.

[36] Assuming $P(x_1)$, the initial density for x_1, is Gaussian and denoting its value with some constant.

In generic notation, we compute

$$\frac{\partial}{\partial\theta}E\left[P\left(\mathbf{x},\mathbf{y}\right)|\mathbf{y}\right], \tag{24}$$

where θ denotes the vector of parameters. In our case, the parameter vector θ is given by $\theta = (A, B, C, Q, R)$.

At this point, we only point out the differences relative to the case without control signal, more details can be found in appendix A.2. Our estimating equations for C and R are the same as in (Ghahramani and Hinton, 1996), who derive expressions for A, C, Q and R without the control signal u_t. For A and Q, the expressions are different, the difference being due to the terms featuring B in equation (23). Appendix A.2 shows the estimation equation for B.

In the EM algorithm, the Kalman filter and smoother procedures are run first, which yields $E\left[P\left(\mathbf{x},\mathbf{y}\right)|\mathbf{y}\right]$. Then, we calculate the new value of θ and rerun the Kalman filter and smoother again. Repeating this procedures till convergence gives the estimate of θ.

5.6 Results

5.6.1 Manufacturing

Table (4) summarizes estimation results from the state space model in equations (17)-(18). Each column corresponds to a different specification. 'Model 1' describes results when micro-aggregated components are not used in the estimation. In the other models, we used micro-aggregated components to estimate trend productivity growth. In 'Model 2', we used weighted input shares to generate micro-aggregated components. In 'Model 3', we used observed value added shares to generate micro-aggregated components. In 'Model 4', we used the fitted values of value added shares and in 'Model 5', we used fitted values of input shares and productivity changes.[37]

Row 2 shows the maximized value of the log-likelihood. It is the largest with micro-aggregated components using input shares ('Model 2') and the smallest with micro-aggregated components using observed value added shares ('Model 3'). Row 3 shows that the EM-algorithm converges fast.[38] The row labeled "ρ" shows the autoregressive coefficient of the productivity-gap. It is equal to 0.42 when micro-aggregated components are not used in the estimation ('Model 1'). It is between 0.33 and 0.35 when micro-aggregated components are used. The row labeled "b" shows the estimated effect of hours-gap on the productivity gap. Without components ('Model 1') it is -0.02. The row labeled "b_4" shows the effect when micro-aggregated components are used in the estimation.

---

[37] We derived fitted values as described earlier and in appendix A.1.

[38] The initial conditions have moderate effect on parameter estimates and likelihoods. Using $(1, .5, 0, -.5, 1)$ as initial conditions for ρ and b in 'Model 1' yielded likelihoods in the interval $\ell\ell \in [-39.88, -35.11]$ and $\widehat{\rho} \in [.31, .42]$ and $\widehat{b} \in [-.04, -.02]$. Similar initial values for ρ and matrix B yielded $\ell\ell \in [-34.02, -33.26]$ in 'Model 2', $\ell\ell \in [-48.86, -42.53]$ in 'Model 3', $\ell\ell \in [-59.55, -39.86]$ in 'Model 4', $\ell\ell \in [-42.50, -39.86]$ in 'Model 5'.

The estimated effect of the micro-aggregated components on trend productivity growth are shown in rows labeled "b_1"-"b_3". The signs of these coefficients depend on how components are generated. For instance, using weighted input shares ('Model 2') yields negative signs for the between and net entry component. When we use observed value-added or fitted value-added shares to generate micro-aggregated components ('Model 3' and 'Model 4'), the signs are positive.

Table 4: Estimation results 1, differences, data scaled by a factor of 100

	'Model 1'	'Model 2'	'Model 3'	'Model 4'	'Model 5'
Intial log-likelihood	-208.47	-44.18	-44.18	-212.1	-44.2
Maximized log-likelihood	-35.11	-33.26	-42.53	-39.86	-39.80
Number of iterations	6	6	16	3	15
ρ	0.42	0.35	0.33	0.25	0.30
b	-0.02	.	.	.	.
b_1	.	0.01	0.10	0.12	0.83
b_2	.	-0.43	0.63	0.43	-0.13
b_3	.	-0.48	0.32	0.11	0.00
b_4	.	0.04	-0.02	-0.02	-0.06

'Model 1': without microcomponents; 'Model 2': input share-weighted microcomponents; 'Model 3': value-added share-weighted microcomponents; 'Model 4': microcomponents using fitted value-added shares; 'Model 5': microcomponents using fitted input shares and productivity changes. Fitted values were derived as in Chapter 1. Parameter constraints in each model are implied by the model in equations (17) and (18): $A_{12}=A_{21}=0$, $B_{14}=B_{21}=B_{22}=B_{23}=0$, $C=(1\ 1)$.

Model selection proves to be critical and cumbersome. If we select a model based on its likelihood, then we would choose 'Model 2.' If we take into account that the definition of aggregate productivity growth is the sum of each of the micro-aggregated components, then our choice might be 'Model 3' or 'Model 4', both of which reflect positive coefficients on the micro-aggregated components.[39]

We invoke visual inspection to get around the problem of model selection. Figures 2, 3, 4, and 5 plot the state variables in equations (17)-(18). Figures 2 and 3 show trend TFP and the productivity gap derived from our Kalman Filter framework, while figures 4 and 5 plot estimates from 1976 to 1995 and forecasts for the years 1996 to 1998. In addition, the upper panel of table 5 contains averages of the published series for manufacturing and averages of the trend growth rates based on the Kalman filter framework with and without the inclusion of the microcomponents. Averages are taken over growth rates for the entire period of data in our sample and are also taken over various subperiods.[40]

We start with trend TFP estimates in figures 2 and 3.[41] For each productivity trend, one observation stands out immediately: all models signal a temporary pickup in productivity

[39] It is not a given that we should expect, *ex ante*, that the coefficients on the microcomponents should be positive. The microcomponent decomposisition relates the change in aggregate productivity to the sum of the changes in each of the microcomponents. In our Kalman Filter framework, we related the trend growth in TFP to a smoothed growth rate in each of the microcomponents.

[40] The subperiods determined by NBER recession dates. In addition the interval starting with 1989 was chosen based on the substantial productivity slowdown in that year.

[41] These estimates are based on a sample from 1974 to 1998.

growth in the early 1980s, with a peak at around $3\frac{3}{4}$ percent. This spike is consistent with the productivity gains seen at the close of the 1981 recession, where the manufacturing sector saw an output spike of 12 percent in 1983. Similarly, (Roberts, 2001) found that trend-TFP-growth peaked slightly below 2 percent for the private non-farm business sector (figures (4) and (8) in his paper). Since we estimate trend growth rates using data on the manufacturing sector only, we deem that our results are in line with his findings.

'Model 1' (figure 2) yields negative trend growth rates in 1979-1981, which does not seem plausible, but is consistent with the dramatic deceleration in manufacturing productivity in the early 1980s. Out of the five models, model 1 estimates the most dramatic productivity slowdown at the end of the 1980s, with trend growth slipping to $\frac{1}{3}$ percent before moving back up.

Table 5: Multifactor Productivity Growth Summary (average percentage point changes) : 1976-1998.

	Manufacturing					
	Published series	Model 1	Model 2	Model 3	Model 4	Model 5
1976-1998	1.3	1.6	1.7	2.5	2.8	2.7
1976-1981	0.7	0.1	0.1	1.3	2.0	1.8
1982-1988	2.2	2.3	2.4	2.8	3.1	2.7
1989-1991	-1.0	0.9	1.6	1.9	1.9	2.3
1992-1998	1.9	2.5	2.5	3.6	3.7	3.6
Forecast						
1996-1998		2.7	2.2	4.6	4.9	4.6

	NonFarm Business					
	Published series	Model 1	Model 2	Model 3	Model 4	Model 5
1976-1998	0.7	0.5	0.4	1.0	1.0	0.9
1976-1981	0.6	0.5	-0.4	0.6	1.0	0.6
1982-1988	0.8	0.5	0.5	0.8	0.8	1.1
1989-1991	-0.1	0.5	0.7	0.8	0.4	0.4
1992-1998	1.0	0.6	0.7	1.6	1.3	1.1
Forecast						
1996-1998		0.5	0.2	2.2	2.9	0.8

Note: All growth rates are trend growth rates except for the published series calculations.

Adding micro-aggregated components changes the level and the contour of the productivity trend estimates. Figure 3 shows that models 3, 4 and 5 yield an upward sloping trend in the mid-1990s and estimate at about $3\frac{1}{2}$ percent TFP-growth at the end of the sample.[42] The acceleration is less clearly captured by Model 2 where we see a growth rate about $2\frac{1}{2}$ percent at the in the 1992-1998 period. The dynamics of the estimated productivity-gap are similar across the models (not shown). All estimated gaps move in sync with NBER-recessions at the beginning of the 1980s and 1990s.[43]

We now move to our forecasts of trend-TFP in the manufacturing sector.[44] Figure 4 shows

[42] Average growth post the 1991-1992 recession.

[43] The NBER defines the following dates for peaks: quarter 1 in 1980, quarter 3 in 1981 and quarter 3 in 1990. The trough dates are: quarter 3 in 1980, quarter 4 in 1982 and quarter 1 in 1991.

[44] Forecasts are based on a sample from 1974 to 1995.

that when micro-aggregated components are not used, the forecast of trend TFP growth remains flat and slightly above 2 percent. This result directly follows from the autoregressive nature of the general linear model. The forecasts are different when micro-aggregated components enter our Kalman filter framework. Figure 5 and table 5 show that models 3, 4 and 5 all result in a 1 percentage point upward shift in trend growth into the range between 4-5 percent, whereas Model 2 yields a flat forecast slightly above 2 percent.[45]

To sum up, three of our four models using micro-aggregate components detect the temporary pickup in trend productivity growth in the mid-1980s and the acceleration starting in the mid-1990s. In terms of forecasting, the same three models that use micro-aggregated components signal further growth in trend for 1996 to 1998, which is valuable information relative to the model without microcomponents. Moreover, the 90 percent confidence bands indicate that the acceleration in trend productivity is significantly different from zero for both trend estimates and the productivity forecasts for models 3, 4, and 5.

5.6.2 Private Nonfarm Business sector

Table (6) summarizes estimation results for the private nonfarm business sector. Similar to the previous section, the column labeled 'Model 1' shows results we obtained without using micro-aggregated components in the estimation. The other columns corresponds to specifications where we used micro-aggregated components generated using firm-level data on manufacturing. In this case, though, we are using the manufacturing micro-aggregated components to inform our estimates and forecast of total nonfarm business multifactor productivity.

The maximized log-likelihood is the largest with micro-aggregated components using input shares ('Model 2') and the smallest with micro-aggregated components using fitted value added shares ('Model 4'). Row 3 shows that, relative to manufacturing, it takes longer to achieve convergence.[46] The estimate of ρ falls in the interval $[-0.04, -0.48]$. The estimated effect of hours-gap on the productivity gap varies between -0.02 and -0.28 (rows labelled "b" and "b4") depending on whether or not micro-aggregated components are used in the estimation. The signs of the coefficients on the micro-aggregated components depend on how micro-aggregated components are generated. For instance, using weighted input shares ('Model 2') yields negative signs for the between and net entry component. When we use observed value-added or fitted value-added shares to generate micro-aggregated components ('Model 3' and 'Model 4'), the

[45] This is due to estimated coefficients in the B matrix. Coefficients of the components are negative for baseline shares ('Model 2'). They are positive, as expected, for observed shares ('Model 3'). The correlations among the components revealed that the within and net entry components are closely correlated between observed and input shares but the between terms are not. The between term with input shares dips in 1981 whereas the between with observed shares peaks in 1982. This suggests there may be measurement error at the beginning of our sample, which is to be explored in future research.

[46] The initial conditions had moderate effect on parameter estimates and likelihoods. Using $(1, .5, 0, -.5, 1)$ as initial conditions for ρ and b in 'Model 1' yielded likelihoods in the interval $[-38.65, -38.03]$ and $\widehat{\rho} \in [-.48, -.36]$ and $\widehat{b} = (-.05, -.02)$. Similar initial values for ρ and matrix B yielded $\ell\ell \in [-46.53, -37.33]$ in 'Model 2', $\ell\ell \in [-41.06, -39.88]$ in 'Model 3', $\ell\ell \in [-47.18, -43.44]$ in 'Model 4', $\ell\ell \in [-48.09, -40.63]$ in 'Model 5'.

signs are positive.

Table 6: Estimation results Private Nonfarm Business sector, differences

	'Model 1'	'Model 2'	'Model 3'	'Model 4'	'Model 5'
Intial log-likelihood	-85.96	-293.09	-234.58	-187.8	-106.0
Maximized log-likelihood	-38.03	-37.33	-39.88	-43.44	-40.63
Number of iterations	29	14	25	4	3
ρ	-0.477	-0.41	-0.24	-0.04	-0.42
b	-0.023	.	.	.	.
b_1	.	0.004	0.09	0.06	-0.20
b_2	.	-0.18	0.37	-0.25	1.13
b_3	.	-0.13	0.30	0.25	0.46
b_4	.	-0.04	-0.04	-0.28	-0.22

'Model 1': without microcomponents; 'Model 2': input share-weighted microcomponents; 'Model 3': value-added share-weighted microcomponents; 'Model 4': microcomponents using fitted value-added shares; 'Model 5': microcomponents using fitted input shares and productivity changes. Fitted values were derived as in Chapter 1. Parameter constraints in each model are implied by the model in equations (17) and (18): A_{12}=A_{21}=0, B_{14}=B_{21}=B_{22}=B_{23}=0, C=(1 1).

Regarding model selection, the same issues emerge as in the previous section, as we rely on visual inspection of figures 6-9 and the average growth rates found in the lower panel of table 5 to judge the value of the microcomponents in our Kalman filter framework. Overall, the results are similar to those reported in the manufacturing forecasting exercises.[47]

First, note that the trend growth in nonfarm business total factor productivity is roughly flat in the model without microcomponents (figure 6 and table 5).[48] Moreover, given the 90 percent confidence bands, the no-microcomponents framework is only significantly different from zero in the mid 1970s. Comparing the average growth rates from 1976 to 1998 in table 5, the growth rate of trend in models 3,4, and 5 are roughly $\frac{1}{2}$ percentage point higher than in models 1 and 2.

In terms of detecting the productivity pickup in the 1990s, models 3 and 4 exhibit a acceleration in trend growth, while trend growth in model 5 is significantly above zero, albeit at more or less at a constant growth rate. This can be seen in the lower panel of table 5, where trend productivity growth is roughly 1 percentage point higher in the 1992-1998 time period in models 3,4, and 5, as opposed to our no-microcomponents framework. That said, growth decelerates somewhat after 1992 in model 5.

Moving to our forecast, figures 8 and 9, display the trend estimates and the forecasts for the years 1995 through 1998. Again, relative to the no-microcomponent framework, the addition of the microcomponents signals an acceleration in productivity growth moving forward in models 3 and 4 as opposed to the flat and indistinguishable from zero forecast in the no-microcomponent framework. It is important to note that the signal extracted from the microcomponent forecasting framework is even more striking given that we are using annual data through 1995 in these exercises.

[47] Also, the estimated dynamics of the productivity-gap do not contradict stylized facts. However, their support is weaker than in Manufacturing.

[48] Forecasts are based on a sample from 1974 to 1995.

Taken altogether, the addition of the microcomponents adds valuable information to the estimation and forecast of trend multifactor productivity in the non-farm business sector. This point is particularly noteworthy given that we are employing data at an annual frequency over a relatively short time frame. That said, the growth rates in trend TFP for the nonfarm business sector are higher than those found in (Roberts, 2001). More fundamentally, though, despite the parsimonious nature of our application effectively captures the upswing in trend multifactor productivity growth in the late 1990s.

6 Conclusion

This paper extends (Bartelsman and Wolf, 2009) to estimate and forecast trend total factor productivity growth in the U.S. We find the addition of micro-aggregated components, i.e., microcomponents, improves the measurement and forecasting of manufacturing and nonfarm business total factor productivity. Further, we are able to extend models used to forecast structural productivity movements to incorporate the information from the plant-level data from the Census of Manufacturers and Annual Survey of Manufactures. Despite the fact that the plant-level dataset only covers the manufacturing sector, our results in forecasting TFP for the nonfarm private business are encouraging.

We first forecast productivity in the manufacturing sector using univariate and multivariate exercises and find that they addition of microcomponents measurably improves forecasts in three year forecast windows. We then build a state-space framework to estimate structural TFP and forecast trend TFP for the years 1996 to 1998. While our results appear stronger for manufacturing than for the private nonfarm business sector, both forecasting exercises capture the productivity acceleration in the second half of the 1990s.

The time-span of our plant-level sample is 25 years, which is sufficient to identify the elasticities of input factors used to calculate total factor productivity. On the other hand, the filtering and smoothing algorithms we use are usually applied to either longer timeseries or data at higher frequencies. In addition, the time-horizon of the micro-data limits the timeseries variation we can exploit using aggregate data. This, in turn, may make it difficult to accurately extract longer trends at the yearly frequency. To be sure, it is remarkable that our models capture the stylized facts of trend productivity growth relatively well. There are multiple avenues for further research, such as using quarterly series with interpolated microcomponents or by attempting to use the LBD to extract microcomponents for the entire economy, or at least several non-manufacturing sectors. In addition, while we have primarily focused on the estimation and forecasting of trend TFP growth, the use of microcomponents should also provide useful information when extending our approach to a full potential output Kalman filter framework.

References

Daron Acemoglu, Philippe Aghion, and Fabrizio Zilibotti. Distance to frontier, selection and economic growth. Discussion Paper 3467, Center for Economic Policy Research, 2002.

Daniel A. Ackerberg, Kevin Caves, and Garth Frazer. Structural indentification of production functions. draft, 2005.

Martin Neil Baily, Charles Hulten, David Campbell, Timothy Bresnahan, and Richard E. Caves. Productivity dynamics in manufacturing plants. *Brookings Papers on Economic Activity*, 1992:187–267, 1992.

Martin Neil Baily, Eric J. Bartelsman, and John Haltiwanger. Labor productivity: Structural change and cyclical dynamics. *The Review of Economics and Statistics*, 83(3):420–433, 2001.

Robert Barro and Xavier Sala–i–Martin. Convergence across states and regions. *Brookings Papers on Economic Activity*, 1:107–182, 1991.

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

Eric J. Bartelsman and Zoltan Wolf. Forecasting productivity using information from firm-level data. *Tinbergen Institute Discussion Papers*, 43(3), 2009.

Eric J. Bartelsman, Randy Becker, and Wayne Gray. The NBER-CES manufacturing industry database. (205), 2000.

Eric J. Bartelsman, John Haltiwanger, and Stefano Scarpetta. Measuring and analyzing coss-country differences in firm dynamics. mimeo, National Bureau of Economic Research, Conference on Research on Income and Wealth, Washington DC., March 2005, 2005.

Eric J. Bartelsman, Jonathan Haskel, and Ralf Martin. Distance to which frontier? Evidence on productivity convergence from international firm-level data. Discussion Paper 7032, Centre for Economic Policy Research, 2008.

Susanto Basu and Miles S. Kimball. Cyclical productivity with unobserved input variation. Working Paper 5915, National Bureau of Economic Research, 1997.

Susanto Basu, John Fernald, and Miles Kimball. Are technology improvements contractionary? International Finance Discussion Papers 625, Board of Governors of the Federal Reserve System, September 1998.

Susanto Basu, Luigi Pascali, Fabio Schiantarelli, and Luis Serven. Productivity, welfare and reallocation: Theory and firm-level evidence. Working Paper 15579, National Bureau of Economic Research, December 2009.

Andrew B. Bernard, J. Bradford Jensen, and Peter K. Schott. Trade costs, firms and productivity. *Journal of Monetary Economics*, 53:917–937, 2006.

Nick Bloom. The impact uncertainty of shocks. *Econometrica*, 77(3):623–685, 2009.

Wesley M. Cohen and Daniel A. Levinthal. Innovation and learning: the two faces of R&D. *The Economic Journal*, 99:569–596, September 1989.

Carol Corrado and Joe Mattey. Capacity utilization. *Journal of Economic Perspectives*, 11:151–67, 1997.

Carol Corrado, Joseph Beaulieu, Eric J. Bartelsman, and Paul Lengermann. Modeling aggregate labor productivity and a disaggregate level: New results for us sectors and industries. mimeo, 2006.

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

Gabriel Fagan, Jerome Henry, and Ricardo Mestre. An Area-Wide Model (AWM) for the Euro Area. *ECB Working Paper*, (42), 2001.

Lucia J. Foster, John Haltiwanger, and C. J. Krizan. Aggregate productivity growth: Lessons from microeconomic evidence. Working Paper 6803, National Bureau of Economic Research, 1997.

Zoubin Ghahramani and Geoffrey E. Hinton. Parameter esimation for linear dynamical systems. Technical Report 96-2, University of Toronto, Department of Computer Science, 1996.

Z. Griliches and H. Regev. Firm productivity in israeli industry 1979-1988. *Journal of Econometrics*, 65:175–203, 1995.

William Gullickson and Michael Harper. Multifactor productivity in U.S. manufacturing, 1949Ű-1983. *Monthly Labor Review*, pages 18–28, October 1987.

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

Charles R. Hulten. Total factor productivity. a short biography. *New Developments in Productivity Analysis*, pages 1–54, 2001.

Ron S. Jarmin and Javier Miranda. The longitudinal business database. Working Paper 02–17, Center for Economic Studies, U.S. Census Bureau, 2002.

Gary Koop. *Bayesian Econometrics*. Wiley and Sons, 2003.

Kenneth N. Kuttner. Estimating potential output as a latent variable. *Journal of Business and Economics Statistics*, 12:361–368, 1994.

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

Stephen D. Oliner, Daniel E. Sichel, and Kevin J. Stiroh. Explaining a productive decade. Finance and Economics Discussion Series 63, Board of Governors of the Federal Reserve System, 2007.

G. Steven Olley and Ariel Pakes. The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297, November 1996.

Amil Petrin and James Levinsohn. Measuring aggregate productivity growth using plant-level data. mimeo, 2010.

Amil Petrin, Kirk White, and Jerome Reiter. The impact of plant-level resource reallocations and technical progress on u.s. macroecnomic growth. Working paper, 2009].

Valerie A. Ramey and Daniel J. Vine. Declining volatility in the u.s. automobile industry. *American Economic Review*, 96:1876–1889, 2006.

John M. Roberts. Estimates of the productivity trend using time-varying parameter techniques. *The B.E. Journal of Macroeconomics*, 1, 2001.

Zoltan Wolf. Forecastability, procyclicality and the granular hypothesis. PhD Dissertation, Chapter 3, VU University Amsterdam, 2011.

Jeffrey M. Wooldridge. On estimating firm-level productivon functions using proxy variables to control for unobservables. *Economics Letters*, 104:112–114, 2009.

Figures 2 and 3: Kalman Filtered Trend Multifactor Productivity and Productivity Gap for the Manufacturing Sector

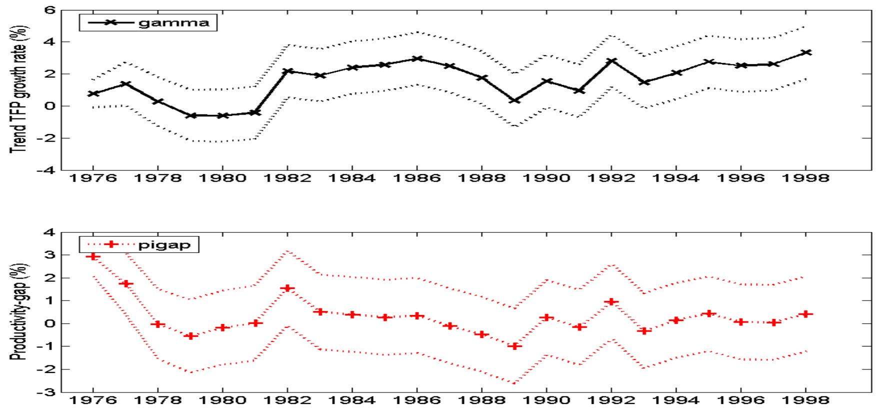

Figure 2: Model 1, without microcomponents, in differences. Dashed lines denote confidence bands with 90% inclusion probability. Trend TFP growth is labeled "gamma", the productivity gap is labeled "pigap".

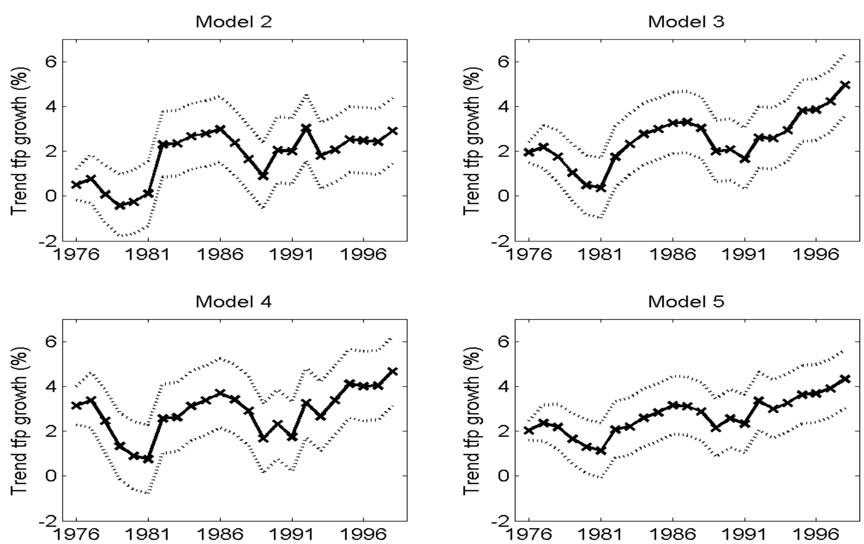

Figure 3: Models 2-5, in differences. Dashed lines denote confidence bands with 90% inclusion probability.

Figures 4 and 5: Kalman Filtered Trend and Trend Forecast for Multifactor Productivity and Productivity Gap for the Manufacturing Sector

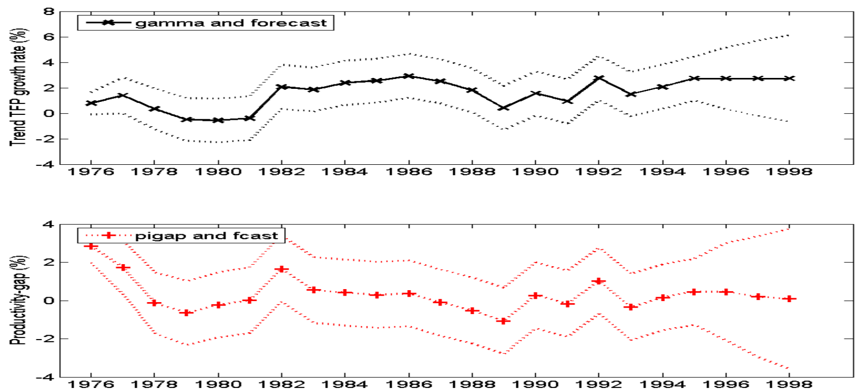

Figure 4: Model 1, without microcomponents, in differences, with forecast period 1996-1998. Dashed lines denote confidence bands with 90% inclusion probability. Trend TFP growth is labeled "gamma", the productivity gap is labeled "pigap".

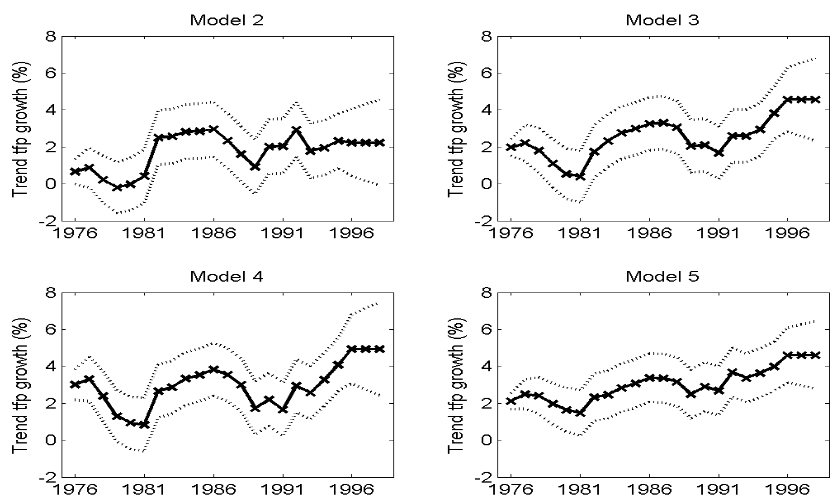

Figure 5: Models 2-5, in differences, forecast period 1996-1998. Dashed lines denote confidence bands with 90% inclusion probability.

Figures 6 and 7: Kalman Filtered Trend Multifactor Productivity and Productivity Gap for the NFB Sector

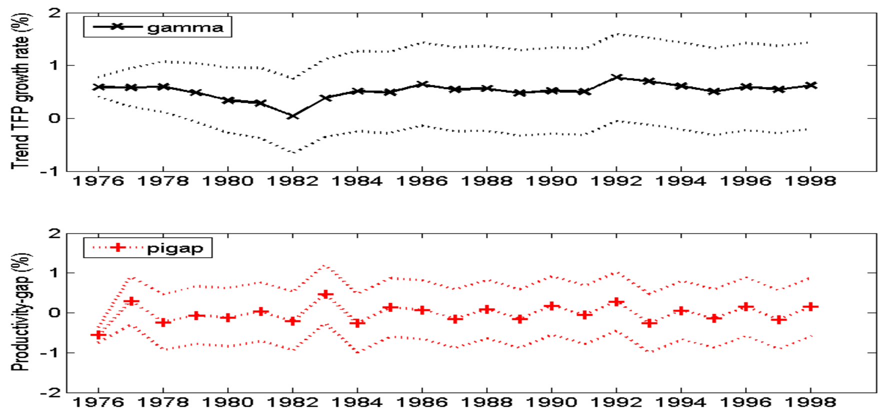

Figure 6: Model 1, without microcomponents, in differences. Dashed lines denote confidence bands with 90% inclusion probability. Trend TFP growth is labeled "gamma", the productivity gap is labeled "pigap".

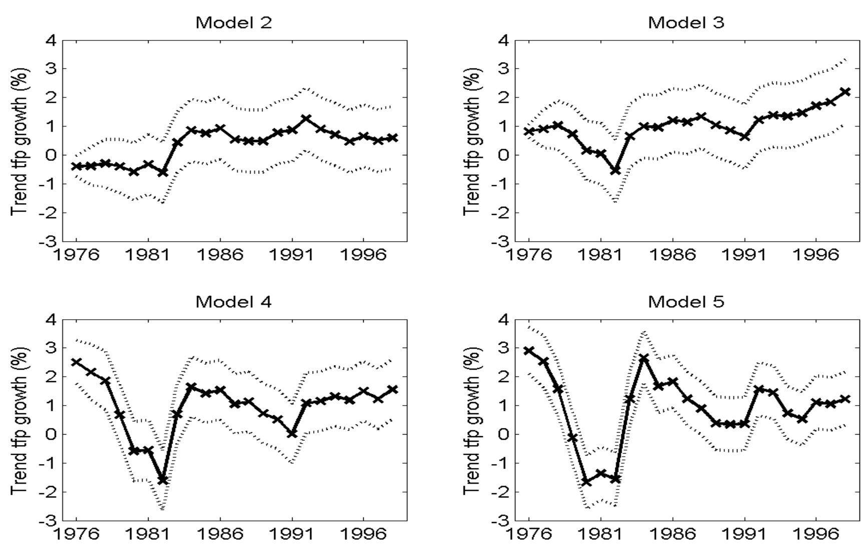

Figure 7: Models 2-5, in differences. Dashed lines denote confidence bands with 90% inclusion probability.

Figures 8 and 9: Kalman Filtered Trend and Trend Forecast for Multifactor Productivity and Productivity Gap for the NFB Sector

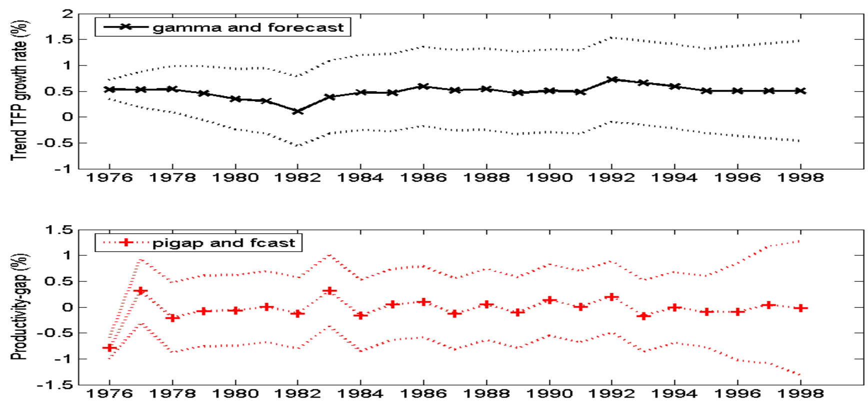

Figure 8: Model 1, without microcomponents, in differences, forecast period 1996-1998. Dashed lines denote confidence bands with 90% inclusion probability. Trend TFP growth is labeled "gamma", the productivity gap is labeled "pigap".

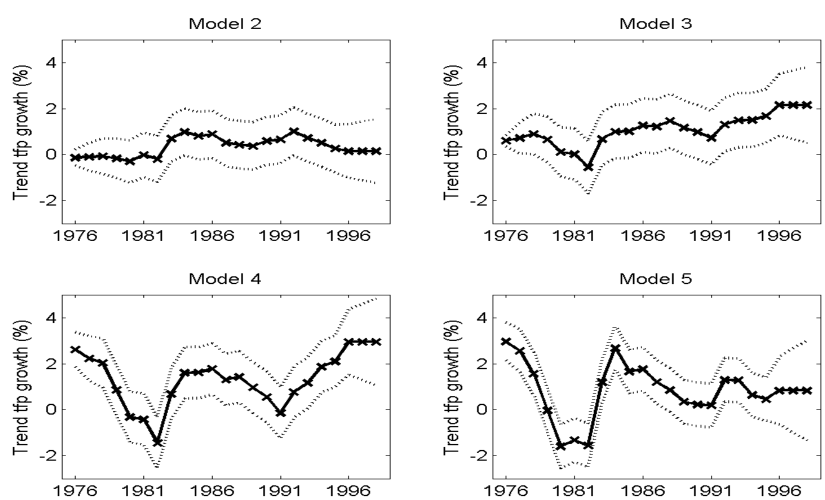

Figure 9: Models 2-5, in differences, forecast period 1996-1998. Dashed lines denote confidence bands with 90% inclusion probability.

Appendix

A.1 Push and pull effects

This section introduces two auxiliary equations. We use them to extract more signal from the distributions of ϕ_{it} and τ_{it}. Our empirical specifications draw on the literature on frontier productivity, both theoretical ((Acemoglu *et al.*, 2002)) and empirical ((Bartelsman *et al.*, 2008)).

The pull-effect is modeled using a firm's position relative to the frontier. The pull-equation posits 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:

$$\Delta\tau_{it} = \beta(\tau_t^F - \tau_{it}) + \eta_{it}, \tag{25}$$

where τ_t^F is frontier productivity and η_{it} is an error term. One can argue for both $\beta < 0$ and $\beta > 0$. If $\beta < 0$, firms closer to the frontier can be thought to have greater absorption capacity[49], so they can take on new technologies and grow faster.[50] If $\beta < 0$, less productive firms tend to grow faster. These firms can be thought as new entrants with relatively high growth potential.

The push equation encapsulates the market selection mechanism whereby less efficient firms are crowded out of the market.[51] Consider the following specification:

$$\phi_{it} = \gamma_0 + \gamma_1(\tau_{it} - \overline{\tau_t}) + \varepsilon_{it}. \tag{26}$$

which implies that, if $\gamma_1 > 0$, higher-than-average productivity firms gain market share, and lower-than-average productivity firms lose market share.

The use of equations (25) and (26) is that the fitted values $\widehat{\Delta\tau_t}$ and $\widehat{\phi_{it}}$ inform about the

[49] One interpretation may be that firms closer to the frontier have larger human capital stock, which is unmeasured and hence it shows up as higher productivity. Also cf. with the "second face" of research in the R&D literature, see, for instance (Cohen and Levinthal, 1989).

[50] The literature calls the $\beta > 0$ case - when less productive firms grow faster - as β-convergence. The concept was introduced by (Barro and Sala–i–Martin, 1991).

[51] Earlier studies found evidence for such reallocation of resources across firms. A recent example is (Bernard *et al.*, 2006).

underlying catch-up components and shares. If the above concepts have a bite in our dataset, $\widehat{\Delta\tau_t}$ and $\widehat{\phi_{it}}$ should add to the forecasting power of the set of components that are computed using these fitted values.

A.1.1 Estimation

Frontier productivity (τ_t^F) is computed as the average of the top decile from a truncated τ-distribution for each industry. We truncate the distribution by eliminating the top and bottom percentiles to avoid the possibly erratic effects on aggregates and also took a moving average of observations to further guard against extreme observations. We calculate τ_t^F using the top decile of this truncated distribution.

The distance variables $(\tau_t^F - \tau_{it})$ and $(\tau_{it} - \overline{\tau}_t)$ in (25) and (26) are clearly correlated with the corresponding error terms η_{it} and ε_{it}, which renders least-squares estimates inconsistent. The endogeneity problem is immediate for equation (25). As for (26), one can argue that a firm gains market share exactly because its productivity increased in the wake of a positive productivity shock.

A simple way to get around the endogeneity problem is to apply an IV estimator. The sample moment condition for $\widehat{\beta}$ 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

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

with $z_{it} = (\tau_{t-2}^F - \tau_{it-2})$. We can also choose lags of Δz_{it} as instruments but that results in the loss of more observations. As for the push-equation (26), rewriting $\phi_{it} = x\gamma' + \varepsilon_{it}$ where $\gamma = [\gamma_0, \gamma_1]$ and $x = [1, \tau_{it} - \overline{\tau}]$. The conventional IV-estimator is given by

$$\widehat{\gamma}_{IV} = \frac{\sum_{i=1}^{N} z_i'\phi_i}{\sum_{i=1}^{N} z_i'x_i} = (Z'X)^{-1} Z'\Phi \text{ with}$$

with $z_i = [(\tau_{i1} - \overline{\tau}_1), ..., (\tau_{iT-1} - \overline{\tau}_{T-1})]'$, $\phi_i = [\phi_{i2}, ..., \phi_{iT}]'$, $Z = [z_1'...z_N']'$, $\Phi = [\phi_1'...\phi_N']'$. The IV-regression results of equations (25) and (26) were in line with our expectations. β appeared

significantly positive at conventional levels in all twelve industries[52,53] The γ's also turned out significantly positive.[54]

A.2 The EM-algorithm

We describe the basic equations of the Expectation Maximization algorithm we use estimating the system matrices of the Kalman filter. The equations for the baseline case without control signal can be found in (Ghahramani and Hinton, 1996).

A.2.1 Model

Our Gaussian model is

$$\begin{aligned} x_{t+1} &= A * x_t + B * u_t + w_t, \\ y_t &= C * x_t + v_t, \\ x(0) &\sim N(init\_x, init\_V) \\ v &\sim N(0, R), \ \ w \sim N(0, Q), \end{aligned}$$

with $x_t \in \mathbb{R}^k$, $y_t \in \mathbb{R}^p$, $A \in \mathbb{R}^{k\times k}$, $B \in \mathbb{R}^{k\times s}$ and $C \in \mathbb{R}^{p\times k}$.

A.2.2 Complete Data Likelihood

The density of an observation conditional on the state is

$$P(y_t|x_t) = \exp\left\{-\frac{1}{2}\left[y_t - Cx_t\right]' R^{-1}\left[y_t - Cx_t\right]\right\}(2\pi)^{-p/2}|R|^{-1/2}, \tag{27}$$

and the density $P(x_t|x_{t-1})$ of the state x_t conditional on its previous value x_{t-1} is

$$\exp\left\{-\frac{1}{2}\left[x_t - Ax_{t-1} - Bu_{t-1}\right]' Q^{-1}\left[x_t - Ax_{t-1} - Bu_{t-1}\right]\right\}(2\pi)^{-k/2}|Q|^{-1/2}. \tag{28}$$

[52] The mean and standard deviation of β across the twelve industries was .039 and .014. Detailed results are available upon request.

[53] Results were very similar in the case of gross output, except that β did not appear significant in the textiles industry. The mean and standard deviation of β across the twelve industries were .032 and .016. Detailed results are available upon request.

[54] The output-based specifications showed more ambiguous relationships: γ_0 were negative in Nonmetallic mineral products and Electrical and optical equipment. Estimates were invariant to whether we measured ϕ_{it} by input- or output-side indicators of firms. That is, using input shares or value added/gross output shares did not affect point estimates at 4-digit precision.

The joint density of the state and the observation vectors is given by

$$P\left(\{\mathbf{x}\},\{\mathbf{y}\}\right)=P\left(x_1\right)\prod_{t=2}^{T}P(x_t|x_{t-1})\prod_{t=1}^{T}P(y_t|x_t). \tag{29}$$

Taking logs and assuming $P\left(x_1\right)$, the initial density for x_1, is Gaussian, and denoting its value with some constant, after some manipulation, we can write the log-likelihood $\log P\left(\{\mathbf{x}\},\{\mathbf{y}\}\right)$ as

$$\begin{aligned}
&const\\
&-\sum_{t=2}^{T}\frac{1}{2}\left[x_t-Ax_{t-1}-Bu_{t-1}\right]'Q^{-1}\left[x_t-Ax_{t-1}-Bu_{t-1}\right]-\frac{T-1}{2}\log|Q|\\
&-\sum_{t=1}^{T}\frac{1}{2}\left[y_t-Cx_t\right]'R^{-1}\left[y_t-Cx_t\right]-\frac{T}{2}\log|R|\\
&-\frac{T(k+p)-k}{2}\log 2\pi.
\end{aligned} \tag{30}$$

The first and last lines in (30) are different from (Ghahramani and Hinton, 1996). The first line includes Bu, the last line has the $-k$ term, which is due to our notation: had we written out the density of x_1, we would have another $(2\pi)^{-k/2}$ term in the likelihood and another $k\log 2\pi$ term in the log-likelihood so the final term above would just be $-\log 2\pi\frac{T(k+p)}{2}$ as in (Ghahramani and Hinton, 1996).

A.2.3 The M-step

Matrix C Denote $E\left[x_t|\mathbf{y}\right]=\hat{x}_t$, $E\left[x_tx_t'|\mathbf{y}\right]=P_t$, and $E\left[x_tx_{t-1}'|\mathbf{y}\right]=\hat{x}_t\hat{x}_{t-1}'=P_{t,t-1}$, implying

$$\begin{aligned}
P_t &= E\left(x_tx_t'|\mathbf{y}\right)\\
&= var(x_t|\mathbf{y})+E\left[x_t|\mathbf{y}\right]E\left[x_t'|\mathbf{y}\right]\\
&P_t^x+\hat{x}_t\hat{x}_t',
\end{aligned}$$

where P_t^x is the smoothed state variance and $\hat{x}_t$ is the smoothed state estimate in time t. Differentiating (30) wrt. C gives

$$\begin{aligned}
&-\sum_{t=1}^{T}2R^{-1}y_t\hat{x}_t'+\sum_{t=1}^{T}2R^{-1}CP_t\\
&=-\sum_{t=1}^{T}2R^{-1}y_t\hat{x}_t'+\sum_{t=1}^{T}2R^{-1}CP_t,
\end{aligned}$$

which should be equal to zero at the optimal C, implying

$$C^*=\left(\sum_{t=1}^{T}y_t\hat{x}_t'\right)\left(\sum_{t=1}^{T}P_t\right)^{-1}. \tag{31}$$

Matrix A Differentiate (30) wrt. A:

$$-\sum_{t=2}^{T} 2Q^{-1}P_{t,t-1} + \sum_{t=2}^{T} 2Q^{-1}AP_{t-1} + \sum_{t=2}^{T} 2Q^{-1}Bu_{t-1}\hat{x}'_{t-1} = 0,$$

which can be rearranged

$$\begin{aligned} A\left(\sum_{t=2}^{T} P_{t-1}\right) &= \sum_{t=2}^{T} P_{t,t-1} - \sum_{t=2}^{T} Bu_{t-1}\hat{x}'_{t-1} \\ A^* &= \left[\sum_{t=2}^{T} P_{t,t-1} - B\sum_{t=2}^{T} P^{ux}_{t-1}\right]\left(\sum_{t=2}^{T} P_{t-1}\right)^{-1}. \end{aligned} \tag{32}$$

In addition to the $P_t, P_{t,t-1}$, and $P_{t-1,t}$ terms, we need to calculate other covariance terms. Note that

$$\begin{aligned} P^{xu}_{t,t-1} &= \hat{x}_t u'_{t-1} = (u_{t-1}\hat{x}'_t)' = \left(P^{ux}_{t-1,t}\right)' \\ P^{xu}_{t-1} &= \hat{x}_{t-1}u'_{t-1} = \left(u_{t-1}\hat{x}'_{t-1}\right)' = \left(P^{ux}_{t-1}\right)' \\ P^{u}_{t-1} &= u_{t-1}u'_{t-1} = \left(u_{t-1}u'_{t-1}\right)' = \left(P^{u}_{t-1}\right)' \\ \sum_{t=2}^{T} P^{xu}_{t,t-1} &= \sum_{t=2}^{T}\left(P^{ux}_{t-1,t}\right)' \\ \sum_{t=2}^{T} P^{xu}_{t-1} &= \sum_{t=2}^{T}\left(P^{ux}_{t-1}\right)'. \end{aligned} \tag{33}$$

Denote the above sums by boldfaced capitals and use the symmetry property

$$\begin{aligned} \mathbf{P}^{xu}_{t,t-1} &= \sum_{t=2}^{T} P^{xu}_{t,t-1} = \sum_{t=2}^{T}\left(P^{ux}_{t-1,t}\right)' = \left(\mathbf{P}^{ux}_{t-1,t}\right)' \\ \mathbf{P}^{xu}_{t-1} &= \sum_{t=2}^{T} P^{xu}_{t-1} = \sum_{t=2}^{T}\left(P^{ux}_{t-1}\right)' = \left(\mathbf{P}^{ux}_{t-1}\right)' \\ \mathbf{P}^{u}_{t-1} &= \sum_{t=2}^{T} P^{u}_{t-1} = \left(\mathbf{P}^{u}_{t-1}\right)' \\ \mathbf{P}_t &= \sum_{t=2}^{T} P_t \\ \mathbf{P}_{t,t-1} &= \sum_{t=2}^{T} P_{t,t-1} \end{aligned} \tag{34}$$

where we used results from equation (33). Using this notation, A^* can be written as

$$A^* = \left[\mathbf{P}_{t,t-1} - B\mathbf{P}^{ux}_{t-1}\right]\left(\mathbf{P}_{t-1}\right)^{-1} \tag{35}$$

Matrix B Differentiate equation (30) wrt. B:

$$
\begin{aligned}
0 = & -2Q^{-1}\left(\sum_{t=2}^{T}\hat{x}_t u'_{t-1}\right) + 2Q^{-1}A\left(\sum_{t=2}^{T}\hat{x}_{t-1}u'_{t-1}\right) \\
& + 2Q^{-1}B\left(\sum_{t=2}^{T}u_{t-1}u'_{t-1}\right),
\end{aligned}
$$

implying

$$
\begin{aligned}
0 &= -\mathbf{P}^{xu}_{t,t-1} + A\mathbf{P}^{xu}_{t-1} + B\mathbf{P}^{u}_{t-1} \\
B^* &= \left(\mathbf{P}^{xu}_{t,t-1} - A\mathbf{P}^{xu}_{t-1}\right)\left(\mathbf{P}^{u}_{t-1}\right)^{-1}.
\end{aligned}
\tag{36}
$$

Reduced forms of matrices A **and** B Equations (35) and (36) can be combined to express A as a function of the variance and covariance terms only. Use from equation (34) that $\mathbf{P}^{xu}_{t-1} = \left(\mathbf{P}^{ux}_{t-1}\right)'$ to write A as

$$
\begin{aligned}
A^*_r = & \left[\mathbf{P}_{t,t-1} - \mathbf{P}^{xu}_{t,t-1}\left(\mathbf{P}^{u}_{t-1}\right)^{-1}\left(\mathbf{P}^{xu}_{t-1}\right)'\right] * \\
& \left[\mathbf{P}_{t-1} - \mathbf{P}^{xu}_{t-1}\left(\mathbf{P}^{u}_{t-1}\right)^{-1}\left(\mathbf{P}^{xu}_{t-1}\right)'\right]^{-1}
\end{aligned}
\tag{37}
$$

and

$$
B^*_r = \left[\mathbf{P}^{xu}_{t,t-1} - A^*_r\mathbf{P}^{xu}_{t-1}\right]\left(\mathbf{P}^{u}_{t-1}\right)^{-1}. \tag{38}
$$

Equations (37) and (38) give the two matrices that determine the dynamics of the state. The variables needed to compute them are

$$
\mathbf{P}_{t,t-1}, \mathbf{P}^{xu}_{t,t-1}, \mathbf{P}^{xu}_{t-1}, \left(\mathbf{P}_{t-1}\right)^{-1}, \left(\mathbf{P}^{u}_{t-1}\right)^{-1}. \tag{39}
$$

Observation covariance matrix R Differentiate equation (30) wrt. R^{-1}:

$$
\begin{aligned}
-\sum_{t=1}^{T}\left(\frac{1}{2}y_t y'_t - C\hat{x}_t y'_t + \frac{1}{2}C\hat{x}_t\hat{x}'_t C'\right) + \frac{T}{2}R = 0 \\
\frac{T}{2}R - \sum_{t=1}^{T}\left(\frac{1}{2}y_t y'_t - C\hat{x}_t y'_t + \frac{1}{2}CP_t C'\right) = 0,
\end{aligned}
$$

which implies

$$
R^* = \frac{1}{T}\sum_{t=1}^{T}\left(y_t y'_t - 2C\hat{x}_t y'_t + CP_t C'\right), \tag{40}
$$

where we used that

$$
x'_t C' R^{-1} y_t = x'_t C'\left(R^{-1}\right)' y_t,
$$

because R^{-1} is symmetric. Then, using (31)

$$
\begin{aligned}
\sum_{t=1}^{T} CP_tC' &= C\mathbf{P}_tC' \\
&= \left(\sum_{t=1}^{T} y_t\hat{x}_t'\right)(\mathbf{P}_t)^{-1}\mathbf{P}_t(\mathbf{P}_t)^{-1\prime}\left(\sum_{t=1}^{T} y_t\hat{x}_t'\right)' \\
&= \left(\sum_{t=1}^{T} y_t\hat{x}_t'\right)(\mathbf{P}_t)^{-1\prime}\left(\sum_{t=1}^{T} (y_t\hat{x}_t')'\right) \\
&= \left(\sum_{t=1}^{T} y_t\hat{x}_t'\right)(\mathbf{P}_t)^{-1\prime}\left(\sum_{t=1}^{T} (\hat{x}_t y_t')\right) \\
&= \left(\sum_{t=1}^{T} y_t\hat{x}_t'\right)C' \\
&= \sum_{t=1}^{T} y_t\hat{x}_t'C'.
\end{aligned}
$$

It is easy to show that $\sum_{t=1}^{T} y_t\hat{x}_t'C' = \sum_{t=1}^{T} C\hat{x}_t y_t'$, so the second and third terms in (40) are

$$
\begin{aligned}
-2C\sum_{t=1}^{T}\hat{x}_t y_t' + \sum_{t=1}^{T} CP_tC' &= -2C\left(\sum_{t=1}^{T}\hat{x}_t y_t'\right) + C\left(\sum_{t=1}^{T}\hat{x}_t y_t'\right) \\
&= C\left(\sum_{t=1}^{T}\hat{x}_t y_t'\right),
\end{aligned}
$$

implying that R^* is given by

$$
R^* = \frac{1}{T}\sum_{t=1}^{T}\left(y_t y_t' - C^*\hat{x}_t y_t'\right). \tag{41}
$$

Sate error covariance matrix Q Compute the partial derivative of equation (30) wrt. Q

$$
\begin{aligned}
&-\frac{T-1}{2}Q - \frac{1}{2}\sum_{t=2}^{T}\{-P_t + 2P_{t-1,t} - AP_{t-1}A' \\
&+ 2BP^{ux}_{t-1,t} - 2BP^{ux}_{t-1}A' - Bu_{t-1}u_{t-1}'B'\},
\end{aligned}
$$

where we used that the partial derivative of $x_{t-1}'A'Q^{-1}x_t = x_{t-1}'A'Q^{-1\prime}x_t$ wrt. Q^{-1} is $Ax_{t-1}x_t' = x_t x_{t-1}'A'$, and that after taking expectations

$$
\begin{aligned}
AP_{t-1,t} &= P_{t,t-1}A' \\
Bu_{t-1}\hat{x}_t' &= \hat{x}_t u_{t-1}'B' \\
Bu_{t-1}\hat{x}_{t-1}'A' &= A\hat{x}_{t-1}u_{t-1}'B'.
\end{aligned} \tag{42}
$$

Rearrange the partial differential

$$
\begin{aligned}
0 &= \frac{T-1}{2}Q - \frac{1}{2}\sum_{t=2}^{T}\{P_t - AP_{t-1,t} - P_{t,t-1}A' + AP_{t-1}A' \\
&- BP^{ux}_{t-1,t} - P^{xu}_{t,t-1}B' + BP^{ux}_{t-1}A' + AP^{xu}_{t-1}B' + BP^{u}_{t-1}B'\} \\
&= \frac{T-1}{2}Q - \frac{1}{2}\sum_{t=2}^{T}\{P_t - AP_{t-1,t} - P_{t,t-1}A' + AP_{t-1}A'\} \\
&+ \frac{1}{2}\sum_{t=2}^{T}\{BP^{ux}_{t-1,t} + P^{xu}_{t,t-1}B' - BP^{ux}_{t-1}A' - AP^{xu}_{t-1}B' - BP^{u}_{t-1}B'\},
\end{aligned}
\tag{43}
$$

and use equation (37) to substitute out A. Rewrite terms in equation (43) as

$$
\begin{aligned}
&- \sum_{t=2}^{T} P_{t,t-1}A' + \sum_{t=2}^{T} AP_{t-1}A' \\
&= -\left(\sum_{t=2}^{T} P_{t,t-1}\right)(\mathbf{P}_{t-1})^{-1\prime}\left[\mathbf{P}'_{t,t-1} - \mathbf{P}^{ux\prime}_{t-1}B'\right] \\
&+ \left[\mathbf{P}_{t,t-1} - B\mathbf{P}^{ux}_{t-1}\right](\mathbf{P}_{t-1})^{-1}\left(\sum_{t=2}^{T} P_{t-1}\right)(\mathbf{P}_{t-1})^{-1\prime}\left[\mathbf{P}'_{t,t-1} - \mathbf{P}^{ux\prime}_{t-1}B'\right] \\
&= -\mathbf{P}_{t,t-1}(\mathbf{P}_{t-1})^{-1\prime}\left[\mathbf{P}'_{t,t-1} - \mathbf{P}^{ux\prime}_{t-1}B'\right] \\
&+ \left[\mathbf{P}_{t,t-1} - B\mathbf{P}^{ux}_{t-1}\right](\mathbf{P}_{t-1})^{-1}\mathbf{P}_{t-1}(\mathbf{P}_{t-1})^{-1\prime}\left[\mathbf{P}'_{t,t-1} - \mathbf{P}^{ux\prime}_{t-1}B'\right] \\
&= -\mathbf{P}_{t,t-1}(\mathbf{P}_{t-1})^{-1\prime}\mathbf{P}'_{t,t-1} + \mathbf{P}_{t,t-1}(\mathbf{P}_{t-1})^{-1\prime}\mathbf{P}^{ux\prime}_{t-1}B' \\
&+ \mathbf{P}_{t,t-1}(\mathbf{P}_{t-1})^{-1}\mathbf{P}_{t-1}(\mathbf{P}_{t-1})^{-1\prime}\mathbf{P}'_{t,t-1} - \mathbf{P}_{t,t-1}(\mathbf{P}_{t-1})^{-1}\mathbf{P}_{t-1}(\mathbf{P}_{t-1})^{-1\prime}\mathbf{P}^{ux\prime}_{t-1}B' \\
&- B\mathbf{P}^{ux}_{t-1}(\mathbf{P}_{t-1})^{-1}\mathbf{P}_{t-1}(\mathbf{P}_{t-1})^{-1\prime}\mathbf{P}'_{t,t-1} + B\mathbf{P}^{ux}_{t-1}(\mathbf{P}_{t-1})^{-1}\mathbf{P}_{t-1}(\mathbf{P}_{t-1})^{-1\prime}\mathbf{P}^{ux\prime}_{t-1}B' \\
&= -\mathbf{P}_{t,t-1}(\mathbf{P}_{t-1})^{-1\prime}\mathbf{P}'_{t,t-1} + \mathbf{P}_{t,t-1}(\mathbf{P}_{t-1})^{-1\prime}\mathbf{P}^{ux\prime}_{t-1}B' \\
&+ \mathbf{P}_{t,t-1}(\mathbf{P}_{t-1})^{-1\prime}\mathbf{P}'_{t,t-1} - \mathbf{P}_{t,t-1}(\mathbf{P}_{t-1})^{-1\prime}\mathbf{P}^{ux\prime}_{t-1}B' \\
&- B\mathbf{P}^{ux}_{t-1}(\mathbf{P}_{t-1})^{-1\prime}\mathbf{P}'_{t,t-1} + B\mathbf{P}^{ux}_{t-1}(\mathbf{P}_{t-1})^{-1\prime}\mathbf{P}^{ux\prime}_{t-1}B' \\
&= -B\mathbf{P}^{ux}_{t-1}(\mathbf{P}_{t-1})^{-1\prime}\mathbf{P}'_{t,t-1} + B\mathbf{P}^{ux}_{t-1}(\mathbf{P}_{t-1})^{-1\prime}\mathbf{P}^{ux\prime}_{t-1}B'.
\end{aligned}
$$

Using the last line above, we have that equation (43) is

$$
\begin{aligned}
&\frac{T-1}{2}Q \\
&- \frac{1}{2}\mathbf{P}_t + \frac{1}{2}A\mathbf{P}_{t-1,t} + \frac{1}{2}B\mathbf{P}^{ux}_{t-1}(\mathbf{P}_{t-1})^{-1\prime}\mathbf{P}'_{t,t-1} - \frac{1}{2}B\mathbf{P}^{ux}_{t-1}(\mathbf{P}_{t-1})^{-1\prime}\mathbf{P}^{ux\prime}_{t-1}B' \\
&+ \frac{1}{2}\sum_{t=2}^{T}\{BP^{ux}_{t-1,t} + P^{xu}_{t,t-1}B' - BP^{ux}_{t-1}A' - AP^{xu}_{t-1}B' - BP^{u}_{t-1}B'\}.
\end{aligned}
\tag{44}
$$

Simplify, collect terms and rearrange to get the estimating equation for Q:

$$
\begin{aligned}
Q^* = \frac{1}{T-1} \{ & \mathbf{P}_t - A \left[\mathbf{P}_{t-1,t} - \mathbf{P}^{xu}_{t-1} B' \right] \\
& - 2B \mathbf{P}^{ux}_{t-1,t} + B \mathbf{P}^{u}_{t-1} B' \}. \qquad (45)
\end{aligned}
$$

Summary of the M-step We have the following results:

$$
\begin{aligned}
C^* &= \left(\sum_{t=1}^{T} y_t \hat{x}'_t \right) \left(\sum_{t=1}^{T} P_t \right)^{-1} = \left(\sum_{t=1}^{T} y_t \hat{x}'_t \right) (\mathbf{P}_t)^{-1}. && (46) \\
R^* &= \frac{1}{T} \sum_{t=1}^{T} (y_t y'_t - C^* \hat{x}_t y'_t) && (47) \\
A^* &= \left[\sum_{t=2}^{T} P_{t,t-1} - B \sum_{t=2}^{T} u_{t-1} \hat{x}'_{t-1} \right] \left(\sum_{t=2}^{T} P_{t-1} \right)^{-1} = \\
& \left[\mathbf{P}_{t,t-1} - \mathbf{P}^{xu}_{t,t-1} \left(\mathbf{P}^{u}_{t-1} \right)^{-1} \left(\mathbf{P}^{xu}_{t-1} \right)' \right] * \\
& \left[\mathbf{P}_{t-1} - \mathbf{P}^{xu}_{t-1} \left(\mathbf{P}^{u}_{t-1} \right)^{-1} \left(\mathbf{P}^{xu}_{t-1} \right)' \right]^{-1} && (48) \\
B^* &= \left(\mathbf{P}^{xu}_{t,t-1} - A^* \mathbf{P}^{xu}_{t-1} \right) \left(\mathbf{P}^{u}_{t-1} \right)^{-1} && (49) \\
Q^* &= \frac{1}{T-1} \{ \mathbf{P}_t - A^* \left[\mathbf{P}_{t-1,t} - \mathbf{P}^{xu}_{t-1} B^{*\prime} \right] - 2B^* \mathbf{P}^{ux}_{t-1,t} + B^* \mathbf{P}^{u}_{t-1} B^{*\prime} \}, && (50)
\end{aligned}
$$

where the $\mathbf{P}$ terms are defined below[55]:

$$
\begin{aligned}
\mathbf{P}_t &= \sum_{t=2}^{T} P_t = \sum_{t=2}^{T} (P^x_t + \hat{x}_t \hat{x}'_t) \\
\mathbf{P}_{t,t-1} &= \sum_{t=2}^{T} P_{t,t-1} = \sum_{t=2}^{T} \left(P^x_{t,t-1} + \hat{x}_t \hat{x}'_{t-1} \right) \\
\mathbf{P}_{t-1} &= \sum_{t=2}^{T} P_{t-1} = \sum_{t=2}^{T} \left(P^x_{t-1} + \hat{x}_{t-1} \hat{x}'_{t-1} \right) \\
\mathbf{P}^{xu}_{t,t-1} &= \sum_{t=2}^{T} P^{xu}_{t,t-1} = \sum_{t=2}^{T} \hat{x}_t u'_{t-1} && (51) \\
\mathbf{P}^{xu}_{t-1} &= \sum_{t=2}^{T} P^{xu}_{t-1} = \sum_{t=2}^{T} \hat{x}_{t-1} u'_{t-1} \\
\mathbf{P}^{u}_{t-1} &= \sum_{t=2}^{T} P^{u}_{t-1} = \sum_{t=2}^{T} u_{t-1} u'_{t-1},
\end{aligned}
$$

where $P^x_{t,t-1}$ denotes the smoothed estimate of the covariance between x_t and x_{t-1} and $P^{xu}_{t,t-1}$

[55] The summation of P_t terms runs $\sum_{t=1}^{T}$ in the equation for C. The other P terms are summed as $\sum_{t=2}^{T}$. See also the Estep for explanation (*gamma* vs *gamma*1).

denotes the smoothed estimate of the covariance between x_{t-1} and u_{t-1}. (Ghahramani and Hinton, 1996) derive the analytical results for the case without control signal. The estimating equations for C and R are the same for the case with control signal. For A and Q, the deviations from the no-control case are due to the terms featuring B and shown in equations (48) and (50). We need to compute $\mathbf{P}_{t,t-1}, \mathbf{P}^{xu}_{t,t-1}, \mathbf{P}^{xu}_{t-1}, (\mathbf{P}_{t-1})^{-1}, \left(\mathbf{P}^{u}_{t-1}\right)^{-1}$ in (51) in order to calculate A, B and then use A and B to calculate the Q covariance matrix.[56] The M-step in iteration i looks like this:

1. Compute C^i and R^i (equations (46), (47))
2. Compute A^i and B^i (equations (48),(49))
3. Compute Q^i (equation (50))
4. Run the Kalman filter and smoother using the new $\theta^i = (A^i, B^i, C^i, Q^i, R^i)$ and use equations in (51) to compute the new $\mathbf{P}$-terms.
5. Go to step 1 or exit if convergence is achieved.

A.2.4 The E-step

The E-step, it involves calculating the expected value of the complete data likelihood[57]. It amounts to running the Kalman filter and smoother and computing the $\mathbf{P}$-terms, which are used in the estimating equations of the M-step.

[56] Note that, as for the other covariance terms, we have that

$$\begin{aligned}
\mathbf{P}_{t,t-1} &= \sum_{t=2}^{T} P_{t,t-1} = \sum_{t=2}^{T} \left(P^{x}_{t,t-1} + \hat{x}_t \hat{x}'_{t-1}\right) \\
&= \sum_{t=2}^{T} \left(P^{x\prime}_{t-1,t} + (\hat{x}_{t-1} \hat{x}'_t)'\right) \\
&= \sum_{t=2}^{T} P'_{t-1,t} = \mathbf{P}'_{t-1,t}
\end{aligned}$$

[57] The joint density of $\mathbf{y}$ and $\mathbf{x}$, conditional on the parameter vector (A, B, C, Q, R).

To understand the code, note that

$$\begin{aligned}
alpha &= \sum_{t=1}^{T} y_t y_t' \\
beta &= \sum_{t=2}^{T} P_{t,t-1} = \mathbf{P}_{t,t-1} \\
gamma &= \sum_{t=1}^{T} P_t = \sum_{t=1}^{T} (P_t^x + \hat{x}_t \hat{x}_t') = \mathbf{P}_t \\
gamma1 &= \sum_{t=2}^{T} P_{t-1} = \sum_{s=1}^{T-1} P_s = \sum_{s=1}^{T-1} (P_s^x + \hat{x}_s \hat{x}_s') \\
&= gamma - (P_T^x + \hat{x}_T \hat{x}_T') = \mathbf{P}_{t-1} \\
gamma2 &= \sum_{t=2}^{T} P_t = gamma - (P_1^x + \hat{x}_1 \hat{x}_1') \\
delta &= \sum_{t=1}^{T} y_t \hat{x}_t'. \\
zeta &= \sum_{t=1}^{T} P_t^u = \sum_{t=1}^{T} u_t u_t' \\
zeta1 &= \sum_{t=2}^{T} P_{t-1}^u = \sum_{s=1}^{T-1} u_s u_s' = zeta - u_T u_T' = \mathbf{P}_{t-1}^u \\
eta &= \sum_{t=1}^{T} P_t^{xu} = \sum_{t=1}^{T} \hat{x}_t u_t' \\
eta1 &= \sum_{t=2}^{T} P_{t-1}^{xu} = \sum_{s=1}^{T-1} \hat{x}_s u_s' = eta - \hat{x}_T u_T' = \mathbf{P}_{t-1}^{xu} \\
theta &= \sum_{t=2}^{T} P_{t,t-1}^{xu} = \sum_{t=1}^{T} \hat{x}_t u_{t-1}' = \mathbf{P}_{t,t-1}^{xu}
\end{aligned}$$