DEEP DYNAMICS*

Nils Gottfries#, Glenn Mickelsson## and Karolina Stadin###

This version, February 12, 2018

Combining micro and macro data, we construct demand-side shocks, which we take to be exogenous for individual firms, and we estimate a reduced form model to see how firms adjust their production, employment, capita stock, and inventories in response to such shocks. Then we estimate the deep structural parameters of a theoretical model by matching empirical impulse-response functions from the estimated reduced-form model. Firms' reactions to demand-side shocks can be well explained by a model where firms have modest market power, face convex adjustment costs and where they can vary utilization at a cost. The stock-out motive helps to explain inventory dynamics.

Keywords: capacity utilization, factor, labor hoarding, labor productivity, inventory holdings, returns to scale, production function, Solow residual

Department of economics, Uppsala University, UCLS, IZA and CESifo, nils.gottfries@nek.uu.se

Department of economics, Uppsala University, glenn.mickelsson@nek.uu.se

Ratio Institute, Stockholm, and UCLS, Uppsala University, karolina.stadin@nek.uu.se

*We have received helpful comments from many colleagues including Martin M. Andreasen, Selva Bahar Baziki, Mikael Carlsson, Daria Finocchiaro, Peter Fredriksson, Georg Graetz, Bertil Holmlund, Per Krusell, Jesper Lindé, Florin Maican, Erik Spector, Ingvar Strid, Karl Wallentin, Eran Yashiv, Erik Öberg and seminar participants at Greater Stockholm Macro Group, Ratio and Uppsala University. Financial support from the Wallander-Hedelius-Browaldh foundation and the Marianne and Marcus Wallenberg Foundation is gratefully acknowledged.

1. Introduction

Understanding how firms react to shocks is important for understanding business cycles and the potential role of policy. Alternative theories provide potential explanations of key stylized facts, such as the pro-cyclicality of investment, labor input, labor productivity and inventory holdings, but there is no consensus on how we should interpret these correlations.

Pro-cyclical labor and factor productivity has been documented in many studies and the Solow residual has been used to measure technology in the real business cycle literature (Prescott, 1986) but many researchers have questioned the interpretation of changes in the Solow residual as technology shocks.[1] Hall (1988) addressed the problem in a new way by considering variations in inputs and output that are related to shocks (military spending, oil price, and the political party of president) which should be uncorrelated with technology. He showed that variations in production that are associated with these shocks are more than proportional to the associated changes in inputs and he interpreted this as evidence of increasing returns to scale. With increasing returns, firms will make losses if price is equal to marginal cost and since firms typically do not make losses, not even in periods of low demand, Hall concluded that firms must have very substantial market power.

An alternative explanation is that there are costly variations in factor utilization (e.g. effort). If we do not properly account for the cost of increasing utilization, the marginal cost will be underestimated and the markup will be overestimated. Burnside, Eichenbaum and Rebelo (1993) found that a model with constant returns to scale, perfect competition, implementation lags in employment and costly variations in effort, can account for a positive correlation between the growth rates of the Solow residual and government expenditure.

Another stylized fact is that inventory holdings are pro-cyclical. If inventories were held solely to smooth production in the face of demand-side shocks, we would expect inventories to decrease in periods of high demand. Hence, some researchers have viewed pro-cyclical inventory investment as an indication that the cost of producing must be low in boom periods, maybe due to technology shocks, increasing returns or positive externalities. Alternatively, a stock-out avoidance motive can explain pro-cyclical inventory holdings in the face of demand-side shocks (Kahn (1987, 1992), Bils and Kahn (2000)). The basic idea is that

[1] Hart and Malley (1999), Baily, Bartelsman Haltiwanger (2001), and Field (2010) document pro-cyclical productivity for different countries and time periods. The literature on the "paradox of short run increasing returns to labor" goes far back; see Fay and Medoff (1985) for early interpretations.

firms need to have stocks of finished goods on the shelves in order to sell and in order to satisfy higher demand, they need to have more goods on the shelves. More recent studies finding evidence in line with the stock-out avoidance theory are Wen (2005) and Kryvtsov and Midrigan (2013).

In this paper, we try to investigate the relevance of some of the theories described above by using firm-level data to estimate firms´ responses to specific demand-side shocks and building a theoretical model that can match those responses. While we do not directly address the question of the relative importance of supply- and demand-side shocks for business cycle fluctuations, we do estimate deep structural parameters that can help us to understand how firms adjust to shocks and these estimates can be used as benchmarks in the construction of macroeconomic models. Our analysis proceeds in four steps:

First, we construct a *demand index* that varies across firms because the share of the goods that go to domestic consumption and investment and various export markets varies across industries and because the share of the firm's production that is sold in the export market varies across firms. This approach is similar to Hall (1988) and Kryvtsov and Midrigan (2013) in that we try to construct measures of demand-side shocks that should be unrelated to technology and cost shocks that affect individual firms or industries.

Second, we try to capture the dynamic responses of firms by estimating a *reduced-form model* of the firm on panel data for manufacturing firms. The empirical model includes production, the capital stock, the number of (full-time equivalent) employees, the inventory stock and the demand index. The endogenous variables depend on their own lags and on the demand index, which we take to be exogenous for the individual firm and we include fixed effects and time dummies in the estimation.[2] As explained in Section 2, the basic idea is that effects of omitted state variables are "mopped up" by lags of the variables that we can observe. We find that firms react strongly to demand-side shocks. Production, capital, and labor increase, but while production responds quickly to the demand shock, registered inputs respond with very substantial lags. This implies positive responses of factor and labor productivity (as they are normally measured) to demand-side shocks. Inventory holdings also increase, but with a smaller lag relative to production.

Third, we construct a *theoretical model* that incorporates many of the explanations of pro-cyclical labor productivity and inventory holdings that have been suggested in the literature. We assume that firms have market power and that they face adjustment costs and

[2] The model is a reduced-form model of the firm in the sense that all variables on the right hand side are assumed to be either predetermined or strictly exogenous relative to the firm's decision variables.

implementation lags in hiring and investment. We allow firms to vary the utilization of both labor and the capital stock at a cost. Furthermore, workers can spend time on activities, such as training, which increase future production but do not contribute to production in the current period. Inventories are needed to prevent stock-outs but they can also be used to smooth production and we take account of the fact that a large part of the inventory stock consists of inputs.[3]

In the fourth step, we run a "horse race" between alternative theories by estimating the *deep parameters* of our model. We follow the approach of Christiano et al. (2005) by choosing the deep parameters so as to match estimated impulse-response functions from our empirical reduced-form model. We use a search algorithm developed by Mickelsson (2016) to search the parameter space for the parameter combination that gives the best match of the impulse-response functions. Standard errors are calculated by bootstrapping, i.e. resampling from the population of firms with replacement and re-estimating the parameters.

We find that our theoretical model can explain the estimated response of the average firm very well. A strong response of production to the demand shock is explained by firms having market power and firm-level demand being very sensitive to the shock as we measure it. Slow adjustment of inputs is mainly explained by convex adjustment costs but implementation lags (time to build) also play a role. Production increases rapidly with the demand shock and most of the short run adjustment is achieved by increasing utilization. According to our estimates, increasing returns to scale in production play a small role.

Inventory holdings are not used to smooth production but instead inventory investments respond positively to demand-side shocks. This "accelerator effect" on inventory investments is explained partly by the stock-out motive, which affects holdings of finished-goods inventories, and partly by the fact that a substantial fraction of the inventory stock consists of inputs and goods in process, which are necessary for production.

As far as we know, this is the first paper to estimate a structural model of the joint dynamics of capital, labor and inventory holdings using micro data. Since these decisions are intimately linked, it makes sense to model them jointly. Our approach to identification follows Hall (1988) in that we try to isolate movements in inputs, output, and inventory holdings, which are caused by specific demand-side shocks, which should be orthogonal to productivity and cost shocks, but we use micro data instead of macro data. In terms of estimation, we follow Christiano, Eichenbaum and Evans (2005) in that we estimate the structural parameters

[3] The two types of inventories are included separately in the theoretical model but they are not distinguished in these data.

by matching impulse-response functions. However, our impulse-response functions are not obtained from a standard vector-autoregressive model but from a reduced-form model of the firm with the demand variable taken to be exogenous relative to the choice variables of the firm.

In our view, there are two advantages of this approach compared to estimation of a structural model with some specific assumptions about the unobserved shocks. The first is that it allows us to remain agnostic about what the unobserved state variables are; we allow the data to speak more freely compared to if we would estimate a tight structural model. The second advantage is that, by comparing the impulse-response functions in the theoretical model to those estimated using the unconstrained reduced-form model, we can see very clearly why certain features of the model are needed to explain the dynamics and why other features cannot, by themselves, explain the data. For example, a model with increasing returns in production, but no variation in utilization, can explain the "excess" response of production compared to labor input at one horizon but not the whole profile of the impulse-response functions (see Section 6).

In the next section, we motivate the method and in Section 3 we present the data and the estimated empirical reduced-form model. A theoretical model of the firm is presented in Section 4 and Section 5 explains how we estimate the structural parameters. The estimated structural model is presented in Section 6 and Section 7 relates our results to previous research. Section 8 concludes.

2 Using a reduced-form model to elicit firms' responses

To study firm dynamics is hard. In order to fully characterize the firm's dynamic optimization problem we would need to observe a large set of exogenous and endogenous state variables that are relevant for the firm's decisions. The problem is that we cannot observe all the relevant state variables and estimation without some of the state variables will lead to biased estimates.[4]

What we do in this study is that we try to represent the relevant set of state variables by current and lagged values of the variables that we can observe. By estimating a reduced-form model of the firm, with demand modeled as a separate stochastic process, we find out how firms respond to demand shocks, which are constructed to be exogenous relative to the

[4] The same argument applies to Euler equation estimation. If, for example, future and past employment help to explain current employment for a given wage, we may interpret this as an indication of adjustment costs, when in fact it is due to some omitted state variable.

firm and the industry. Aggregate state variables are "mopped up" by time dummies. By matching firm's responses to observed exogenous shocks, we obtain estimates of the parameters in our theoretical model.

To see how this might work (or not work) consider a standard model of a firm with quadratic adjustment costs related to changes in labor and capital. The firm faces a downward-sloping demand curve and production is given by the production function: $y_t = (1-\alpha)n_t + \alpha k_{t-1} + a_t$, where y_t is production, n_t is employment and k_t is the capital stock at the end of the period *t* and a_t is a productivity shock. (All variables are logs and firm-specific but we omit the firm index here.) Although we can observe the labor share, we view α as an unknown parameter because the markup is unknown. Assume that there are two exogenous state variables that matter for the firm: a demand shifter d_t, which we can observe, and firm-specific productivity a_t which we cannot observe. Assume that the logs of the shocks follow AR(1) processes: $d_t = \rho_d d_{t-1} + \varepsilon_{dt}$. and $a_t = \rho_a a_{t-1} + \varepsilon_{at}$ where $E(\varepsilon_{dt}\varepsilon_{at}) = 0$.

The approximate solution to the firm's dynamic optimization problem consists of two log-linear decision rules relating current employment and the capital stock at the end of the period to the initial levels of capital and employment as well as demand and the level of productivity. We add white noise shocks to the decision rules, which we can think of as mistakes:

$$n_t = b_{11}n_{t-1} + b_{12}k_{t-1} + b_{13}d_t + b_{14}a_t + \varepsilon_{nt} \tag{1}$$

$$k_t = b_{21}n_{t-1} + b_{22}k_{t-1} + b_{23}d_t + b_{24}a_t + \varepsilon_{kt}\,. \tag{2}$$

Now, we cannot estimate these decision rules because we do not observe productivity, but we can use the equation for the productivity process to substitute for current productivity and then the production function in period t-1 to substitute for lagged productivity. Doing this, we get a reduced-form model with shocks that are serially uncorrelated and exogenous demand as a "driving force":

$$n_t = b_{11}n_{t-1} + b_{12}k_{t-1} + b_{13}d_t + b_{14}\rho_a\left(y_{t-1} - (1-\alpha)n_{t-1} - \alpha k_{t-2}\right) + b_{14}\varepsilon_{a,t} + \varepsilon_{n,t} \tag{3}$$

$$k_t = b_{21}n_{t-1} + b_{22}k_{t-1} + b_{23}d_t + b_{24}\rho_a\left(y_{t-1} - (1-\alpha)n_{t-1} - \alpha k_{t-2}\right) + b_{24}\varepsilon_{a,t} + \varepsilon_{k,t} \tag{4}$$

$$\begin{aligned} y_t &= (1-\alpha)\left[b_{11}n_{t-1} + b_{12}k_{t-1} + b_{13}d_t + \varepsilon_{nt}\right] + \alpha k_{t-1} \\ &\quad + \left(1 + (1-\alpha)b_{14}\right)\left[\rho_a\left(y_{t-1} - (1-\alpha)n_{t-1} - \alpha k_{t-2}\right) + \varepsilon_{at}\right] \end{aligned} \tag{5}$$

$$d_t = \rho_d d_{t-1} + \varepsilon_{dt}\,. \tag{6}$$

Effectively, we have "mopped up" the effect of the initial firm-specific productivity level by including the lag of production and capital on the right hand side. Note that the coefficient relating current production to lagged production $\left(1+\left(1-\alpha\right)b_{14}\right)\rho_a$ reflects the autoregressive character of the productivity shock *and* the indirect effect of productivity on hiring. If productivity follows an AR(2) process, we can account for this in the same way by including additional lags of production, capital and labor input.

In general, there are many unobserved state variables so linear combinations of observed state variables will be imperfect representations of the unobserved state variables – the "mopping up" will be less than perfect – but we can hope that our reduced-form model captures firms' dynamic responses to the demand shocks in a rough way.

An alternative would be to estimate a standard vector-autoregressive model with the endogenous variables and then use d_t as an instrument for the shocks (see Gertler and Karadi (2015), Ramey (2017)). The advantage of including the exogenous variable d_t explicitly in the system is that demand shocks are not mixed up with other shocks.

The approach to production function estimation proposed by Olley and Pakes (1996) is based on the assumptions that there is only one autoregressive shock variable and there are no adjustment costs for labor, so the decision rule for capital (2) can be inverted and used to eliminate the unobserved shock variable from the system; see Ackerberg, Caves and Frazer (2015).[5] The above model is more general in that there are several shocks and labor is a state variable. In general, however, there will be an imperfect mapping from the unobserved state variables to linear combinations of the observed variables, so the reduced form model should be seen as an approximation.

3. Data and empirical model

In this section we present the firm-level data and the construction of the demand index followed by the presentation of the empirical model and the estimated impulse-response functions.

[5] Levinsohn and Petrin (2003) use inputs instead of capital but their approach builds on the same assumptions.

3.1. Firm-level data

The firm-level panel consists of yearly data for all firms in Sweden. As described below, our main sample consists of firms with at least ten employees in the manufacturing sector 1996-2008.

Firms merge, split and buy plants from each other and it is not obvious when a firm is different enough that it should be regarded as a new firm. In this study we are interested in how established firms react to changes in product demand, so we want to diminish the noise caused by firms merging or buying and selling establishments. We therefore use the FAD units from Statistics Sweden as firm identities. The FAD-units are based on the legal organizational numbers but the FAD number changes if there are large mergers or splits affecting more than 50 percent of the workforce even though the legal organizational number is still the same.[6] When we say "firm" below we refer to the FAD identity.

Real production (*Yr*) is the value of the firm's total production deflated by the producer price for the industry.[7] As a robustness check, we instead use real value added deflated by the value added deflator for the industry (*VAr*) to measure production. In our theory (Section 4) we assume that a fixed quantity of intermediate goods is required to produce one unit of the good which implies that total production and value added should respond in the same way to a shock. As we will see, this is roughly what we find empirically.

The real inventory stock (*Zr*) is the value of the firm's inventories at the end of the year deflated by the producer price for the industry and *N* is the number of employees. The latter is measured in "full-year equivalents" but since we do not have hours data, some of the variation in production per worker will probably reflect variations in overtime.

The real capital stock (*Kr*) consists of machines and buildings. In the firm-level panel data we have the firms' book values of buildings and machinery but generous depreciation allowances imply that the book values of these stocks are much lower than their economic value. With a too low value of the stock of capital we would exaggerate the volatility of the capital stock measured as log changes. For this reason, we tried to construct a better measure of the real capital that a firm has. We did this in three steps: First we obtained indsutry-level estimates of capital stocks and also book values from Statistics Sweden. Using these data for years 2000-2005 we calculated an *average* ratio of book value to economic value on the industry level (2-digit SNI92) for buildings and machines separately. This ratio was then used

[6] Further information on the definitions can be found in the document "Företagens och arbetsställenas dynamik (FAD)" from Statistics Sweden.

[7] Firm-level prices are only available for a subset of firms. We do not use firm-level prices because we want to include as many firms as possible in our estimation and the focus here is not on price-setting.

to scale up the book values of buildings and machines for the first year that a firm appears in the sample. Adding buildings and machines together and dividing by a price index for investments, we express the first-year capital stock in prices of year 2000. Finally, we calculated capital stocks for subsequent years by subtracting depreciation and adding investments in machines and buildings deflated by the investment price index. This was repeated for each year that the firm appears in the sample. We set the depreciation rate of capital to 11 percent based on a weighted average of the depreciation rates for buildings and machines used by SCB. The resulting capital stocks are more than twice as large as the book values and the implied log changes are correspondingly smaller.

3.2 Sample selection

The roles of capital, labor and inventories and the organization of markets vary a great deal across sectors. In this study, we are interested in profit maximizing firms, which produce differentiated products using labor and capital and which have substantial inventory stocks consisting primarily of goods they have produced, inputs, and goods in process. For this reason, we chose to study firms in manufacturing (industries 15-36 according to SNI 92). We need export shares for firms to calculate the demand index and these are not available for firms with less than 10 employees, so we include only firms that have at least ten employees in all their years of existence in the data. Publicly owned firms are dropped because they may have other objectives than privately owned firms. With these exclusions we have a sample of 44-55 thousand observations depending on what variable we consider. *Table 1* shows some descriptive statistics for this sample. The first columns show the statistics for the levels of the variables and the latter columns show some ratios.

In order to create a sample of reasonably homogenous firms and also to deal with measurement errors, we exclude firms which *in some period* had “extreme” levels of production per worker, inventory stock relative to production or capital stock relative to production. With one exception “extreme” is defined as being in the bottom or top 5 percent of the sample of all firm-year observations. The exception is the lower limit for the inventory ratio, which we set to the bottom 25 percent since a relatively big share of the firms have very small inventory stocks. These cutoff limits can be seen in *Table 1*.[8]

[8] Log changes can be very large if stocks are low and we suspect that firms with very large inventory stocks may be involved in extensive trading in addition to producing and storing their own products. Firms with very large capital stocks may be involved in property investment.

Table 1. Descriptive statistics for firm-year observations, manufacturing

Total sample	**Yr**	**N**	**Zr**	**Kr**	**Yr/N**	**Zr/Yr**	**Kr/Yr**	**VAr/N**	**Zr/VAr**	**Kr/VAr**	**VA/Y**	**Kb/Y**
Mean	230000	113	29700	104000	1456	0.1471	0.4428	506	0.5023	1.1649	0.3760	0.2116
Std. d.	1510000	512	159000	727000	1126	0.1953	2.8488	325	10.2011	25.4607	3.7632	0.9830
1%	5804	10	0	65	380	0	0.0040	108	0	0	0.0605	0
5%	8862	12	123	966	***540***	0.0071	***0.0425***	***252***	0.0116	***0.1118***	0.1620	0.0109
25%	18500	18	1556	5013	827	***0.0635***	0.1680	359	***0.1554***	0.4414	0.2936	0.0603
50%	*40600*	*32*	*4952*	*13900*	*1149*	*0.1227*	*0.3305*	*447*	*0.3317*	*0.8746*	*0.3922*	*0.1485*
75%	112000	73	15600	41500	1714	0.1923	0.5730	576	0.5660	1.4813	0.4934	0.2778
95%	699000	365	89200	274000	***3346***	***0.3643***	***1.1454***	***945***	***1.2439***	***3.0839***	0.6477	0.5903
99%	3040000	1208	394000	1610000	5517	0.6057	1.9851	1571	2.3665	5.9384	0.7862	1.0147
Observations	49289	54818	54035	49156	49289	49286	43718	50909	54817	45783	49994	49994

Note: Full, unbalanced panel with all private firms with at least 10 employees in Sweden all their years of existence. Industries included are SNI 15-36, and years included are 1996-2008. X% denotes the Xth percentile. Yr is real production, N is full-time equivalent employees. Zr is real inventory stock, Kr is the real capital stock and VAr is real value added. Real values are thousands SEK in prices of year 2000. PPI for the two-digit sector is used to deflate Y and Z and the value added deflator is used to calculate real value added. The calculation of the real capital stock is described in the text. Kb/Y is nominal book value of capital stock relative to nominal value of production. Boldface numbers are the limits used to delineate the main estimation sample.

Note in *Table 1* that the median ratio of (owned) real capital to total real output (0.33) is more than twice as large as the median book value of capital relative to production, which is (0.15). This is due to depreciation being much higher in the accounting than the estimated economic depreciation rate (11 percent). Value added is about 39 percent of total production (output) in this sample and the median ratio of real capital to value added is 87 percent. This may strike readers as a low value but a substantial fraction of the capital that firms use is rented. Firms often lease cars, trucks and other types of equipment and they may rent the buildings where they operate . Rented capital is treated as a flexible input in our theoretical model.

The resulting sample has 11306 observations that can be used for estimation. In the baseline estimation we impose one further restriction: we include only firms for which we have data the whole period (13 years), reducing the number of observations that we can use in the baseline estimation to 8143. We do this to reduce the "Nickell-bias" in the estimation with fixed effects.[9] If we go in the opposite direction and exclude only "extreme" firm-year *observations* and not the whole time series for the firm with the extreme observation, we get an unbalanced panel with 17179 observations. We report estimated impulse-response functions estimated on the baseline sample and alternative samples below.

3.3 The firm-specific demand index

We construct a firm-specific demand index, $D_{i,t}$, as a weighted average of domestic and international demand for the relevant industry using the firm's export share. The demand index is constructed so as to be as exogenous as possible to the firm and the industry by using only data for components of aggregate demand, data for foreign demand and weights that do not vary over time.[10] To motivate the demand index, let us consider an economy where goods are produced in J different sectors indexed j and used for consumption and investment. Let aggregate investment be a Cobb-Douglas aggregate of goods used for investment and produced in different sectors and where the latter are CES aggregates of goods produced by different firms:

[9] The estimation method is OLS. Nickell bias means that the estimated coefficient for the first lag of the dependent variable tends to be underestimated since some of the variation is instead picked up by the firm fixed effects when there are few observations in the time dimension. We tried to do diff-GMM (Arellano-Bond) estimation, but we were unable to find an instrument set which is both valid and enough relevant to give good identification.

[10] Similar approaches has been used by Lundin et al. (2009), Carlsson et al. (2013), and Eriksson & Stadin (2017). The demand index used here was constructed by Stadin (2015).

$$I=\prod_{j=1}^{J} I_j^{\theta_j^I} \qquad \text{where} \qquad I_j=\left(\sum_{i\in j}\left(Q_i^I\right)^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}}.$$

Q_i^I denotes the amount of goods produced by firm i and used for investment. Investors minimize the cost of a given investment *I* subject to these constraints. Maximizing

$$L=-\sum_{i=1}^{N} P_i D_i^I+\lambda_I\left(\prod_{j=1}^{J} I_j^{\theta_j^I}-I\right)+\sum_{j=1}^{J}\lambda_j\left(\left(\sum_{i\in j}\left(Q_i^I\right)^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}}-I_j\right)$$

with respect to I_j and Q_i^I we get first order conditions $\lambda_j I_j=\theta_j^I\lambda_I I$ and $\mathrm{Q}_i^I=\left(P_i/\lambda_j\right)^{-\eta} I_j$ where λ_I is the relevant price index of aggregate investment, $\lambda_I I$ is total investment expenditure and λ_j is the relevant price index of sector j. Applying the same reasoning to aggregate consumption we get total demand for goods produced by firm *i* in sector *j*:

$$\hat{D}_i=Q_i^I+Q_i^C=\frac{\theta_j^I\lambda_I I}{\lambda_j}\left(\frac{P_i}{\lambda_j}\right)^{-\eta}+\frac{\theta_j^C\lambda_I C}{\lambda_j}\left(\frac{P_i}{\lambda_j}\right)^{-\eta}=\left(\theta_j^I\lambda_I I+\theta_j^C\lambda_C C\right)P_i^{-\eta}\lambda_j^{\eta-1}.$$

Taking logs and log-linearizing with respect to *I* and *C* we get

$$\begin{aligned}
\ln\hat{D}_i &=\ln\left(\theta_j^I\lambda_I e^{\ln I}+\theta_j^C\lambda_C e^{\ln C}\right)-\eta\ln P_i+\left(\eta-1\right)\ln\lambda_j\\
&\approx\ln\bar{D}_i+\frac{\theta_j^I\lambda_I\bar{I}\left(\ln\mathrm{I}-\ln\bar{\mathrm{I}}\right)+\theta_j^C\lambda_C\bar{C}\left(\ln C-\ln\bar{C}\right)}{\theta_j^I\lambda_I\bar{I}+\theta_j^C\lambda_C\bar{C}}-\eta\left(\ln P_i-\ln\bar{P}_i\right)\\
&=\ln\bar{D}_i+\frac{\lambda_j\bar{I}_j}{\lambda_j\bar{I}_j+\lambda_C\bar{C}_j}\left(\ln\mathrm{I}-\ln\bar{\mathrm{I}}\right)+\frac{\lambda_C\bar{C}_j}{\lambda_j\bar{I}_j+\lambda_C\bar{C}_j}\left(\ln C-\ln\bar{C}\right)-\eta\left(\ln P_i-\ln\bar{P}_i\right)
\end{aligned}$$

where bars denote steady state values. We see that the weights are the steady state shares of production in sector j that are used for investment and consumption. The same logic can be applied to sales in different countries and based on this reasoning we construct the demand index:

$$\ln D_{i,t}=\left(1-\delta_i\right)\left[\phi_j^C\ln C_t+\phi_j^G\ln G_t+\phi_j^I\ln I_t+\left(1-\phi_j^C-\phi_j^G-\phi_j^I\right)\ln EX_t\right]+\delta_i\left(\sum_m\omega_{j,m}\ln Y_{j,m,t}^F\right). \quad (7)$$

The subscript *i* denotes the firm, *j* denotes the industry, *t* denotes the year and *m* denotes the country. The weight, δ_i, is the mean export share for the firm over the sample and the weights ϕ_j^C,ϕ_j^G and ϕ_j^I are industry-specific shares calculated on the two-digit level (SNI92) using input-output tables from Statistics Sweden for 2005. The weights are kept fixed over time to make the demand variable as exogenous as possible. ϕ_j^C is private domestic consumption as a share final demand excluding direct exports for the relevant industry and ϕ_j^G and ϕ_j^I are the corresponding shares for public consumption and investment. The remaining share, $1-\phi_j^C-\phi_j^G-\phi_j^I$, is the share of final demand excluding direct exports going *indirectly* to exports for the relevant industry, that is, the share used as intermediate inputs into domestic products that are eventually exported. C_t is real private consumption, G_t is real public consumption, I_t is real gross fixed investment, and EX_t is real exports; all are aggregate values in fixed prices from Statistics Sweden's table for the gross national product from the user side.

The weight $\omega_{j,m}$ is the share of industry *j's* direct exports that goes to country *m*. For some industries there are no data and for those industries, the export share is set to zero. Included export countries are Sweden's main trading partners Germany, Norway, the United Kingdom, Denmark, Finland, the USA, France, the Netherlands, Belgium, Italy, and Spain. The variable $Y_{j,m,t}^F$ is real value added for industry j in country m from the OECD STAN database and it is meant to capture demand in country m for goods produced by sector j.

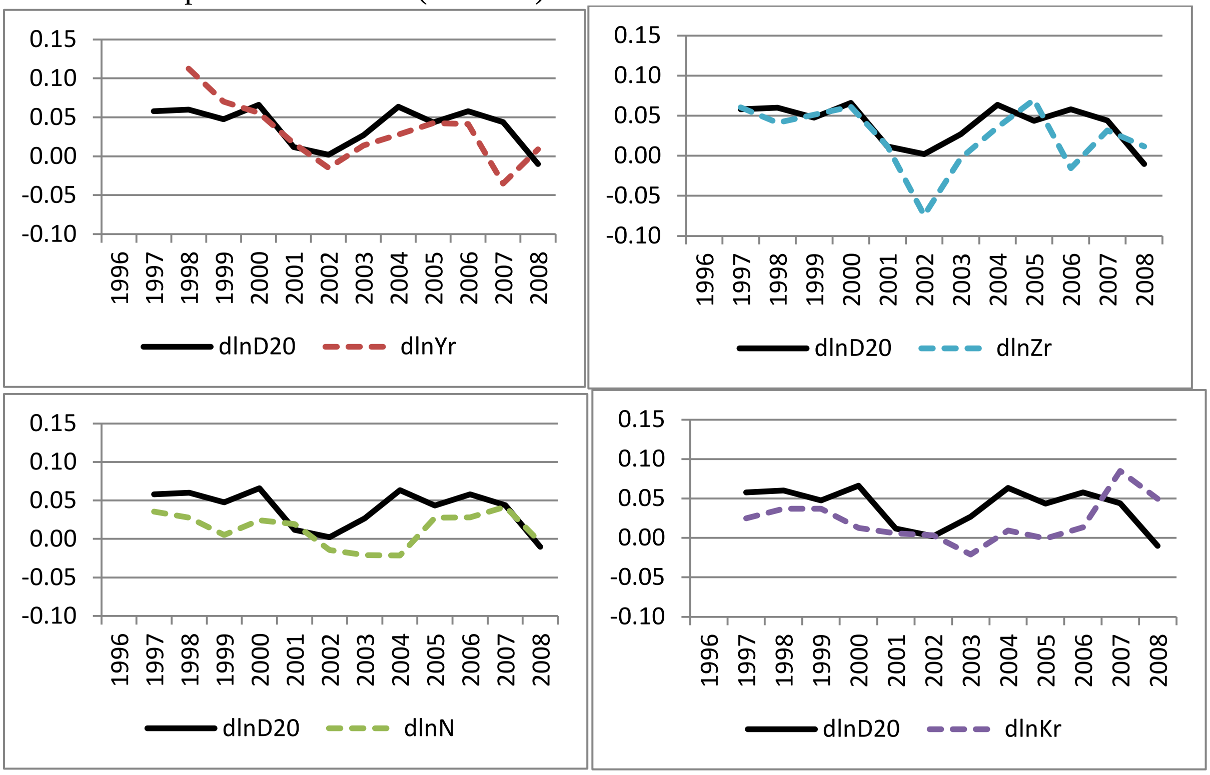
So what types of shocks do the demand variables for different firms represent? Since we have time dummies in our estimation, they will not represent fluctuations in demand, that are common to all firms, but to some extent, they will still reflect the business cycle because investment varies more than consumption over the cycle. Furthermore, they will reflect structural changes in the composition of domestic demand and differences between economic developments in Sweden and foreign markets, which affect firms differently depending on their presence in the export markets.

In order not to introduce spurious correlations due to industry-specific shocks, we do not use industry-specific prices or quantities to construct the demand index. To the extent that industry prices respond to industry demand, we can see the effects of demand shocks as effects of the exogenous shifts in industry demand *and* the industry price responses; both should increase the demand for goods produced by an individual firm.[11]

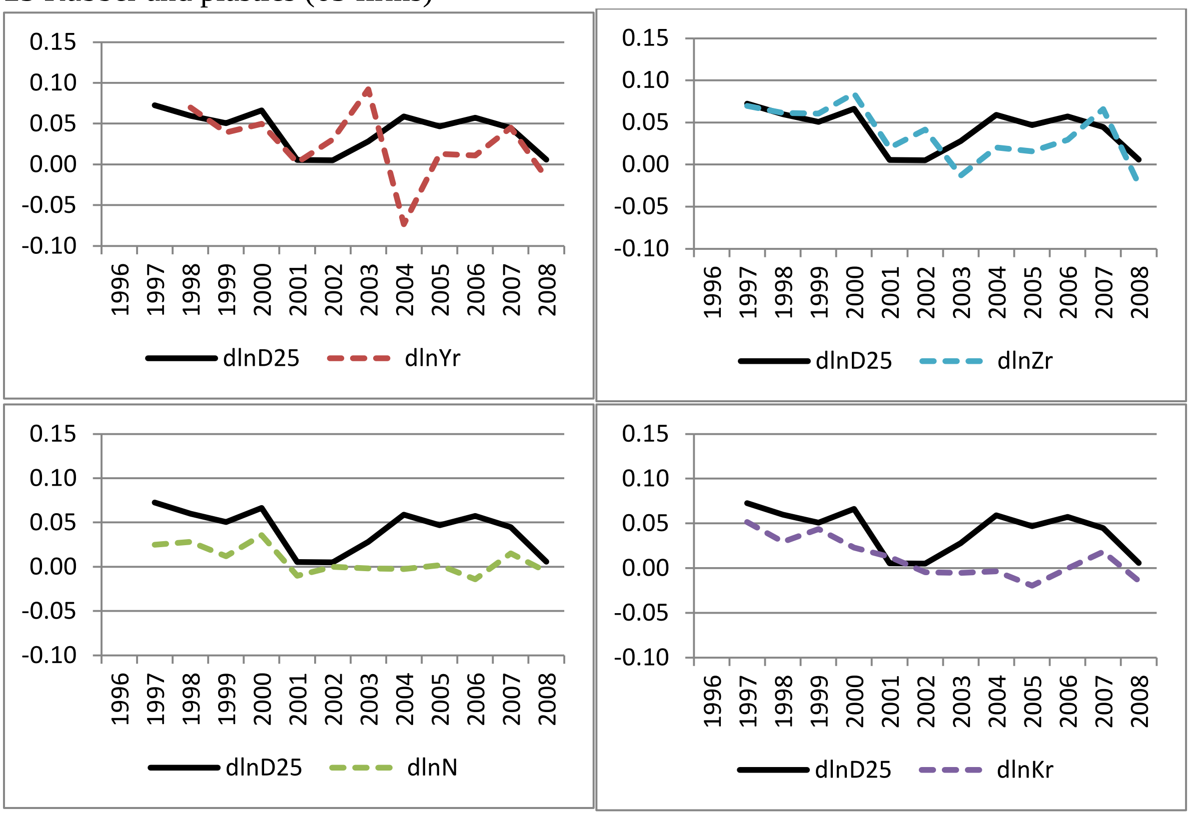
[11] Potential spurious correlation due to variations in the relative price between investment and consumption is discussed below.

Figure 1. Log changes aggregated to the two-digit industry level for 4 industries with the largest number of firms, balanced panel

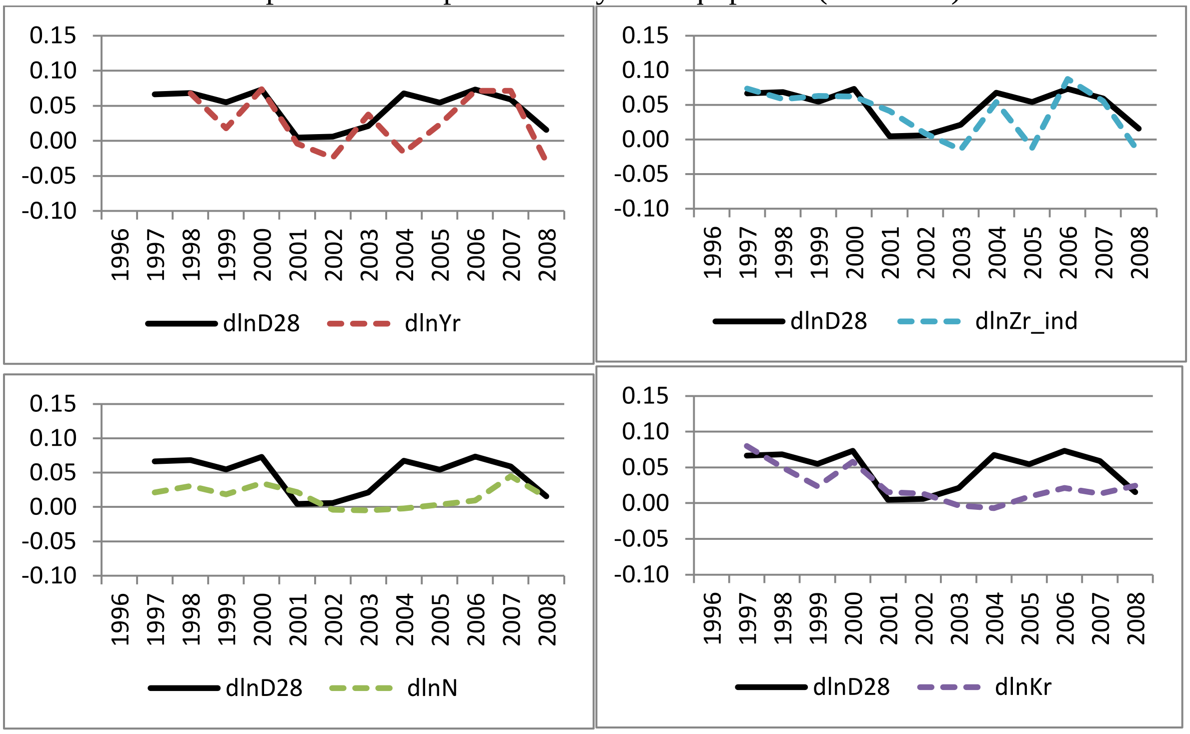
20 Wood and products of wood (53 firms)



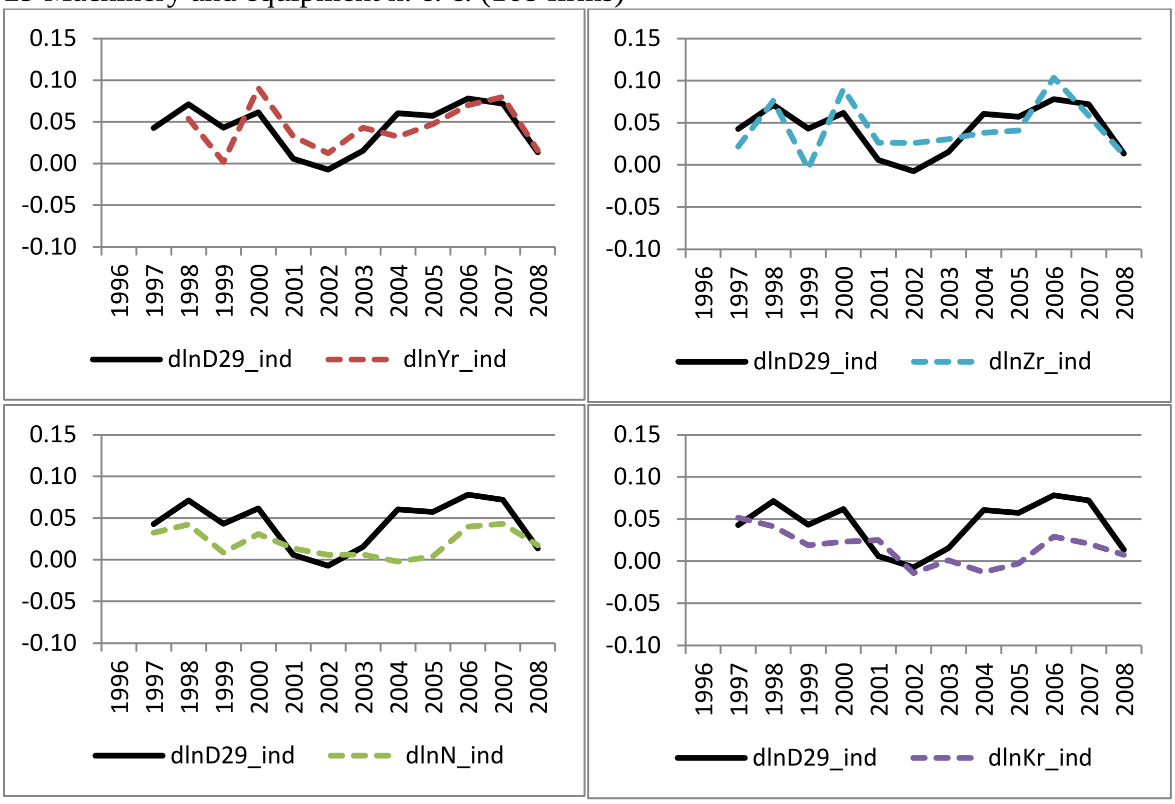
25 Rubber and plastics (65 firms)



28 Fabricated metal products except machinery and equipment (189 firms)



29 Machinery and equipment n. e. c. (168 firms)



For a detailed list of the number of firms per industry, see the Appendix *Table A1*.[12] *Figure 1* illustrates the data for the four industries with the largest number of firms in the balanced panel. For each year, we have taken the log changes for the firms and calculated the average for the industry. In most cases, production and inventories co-vary strongly with the demand index with little or no lag. The changes in the capital stock and the number of workers also co-vary with demand but with a substantial lag.

3.4. Empirical model and identification

To capture the effect of demand shocks on real production (Yr), the real capital stock (Kr), employment (N) and real inventory holdings (Zr), we estimate a reduced-form model with two lags of the endogenous variables and the firm-specific demand index as an exogenous variable:

$$\begin{aligned}\mathrm{lnYr}_{i,t} &= \beta_1^Y \mathrm{lnYr}_{i,t-1} + \beta_2^Y \mathrm{lnYr}_{i,t-2} + \beta_3^Y \mathrm{lnN}_{i,t-1} + \beta_4^Y \mathrm{lnN}_{i,t-2} + \beta_5^Y \mathrm{lnKr}_{i,t-1} + \beta_6^Y \mathrm{lnKr}_{i,t-2} \\ &\quad + \beta_7^Y \mathrm{lnZr}_{i,t-1} + \beta_8^Y \mathrm{lnZr}_{i,t-2} + \beta_9^Y \mathrm{lnD}_{i,t} + \beta_{10}^Y \mathrm{lnD}_{i,t-1} + \varepsilon_{i,t}^Y\end{aligned} \tag{8}$$

$$\begin{aligned}\mathrm{lnN}_{i,t} &= \beta_1^N \mathrm{lnYr}_{i,t-1} + \beta_2^N \mathrm{lnYr}_{i,t-2} + \beta_3^N \mathrm{lnN}_{i,t-1} + \beta_4^N \mathrm{lnN}_{i,t-2} + \beta_5^N \mathrm{lnKr}_{i,t-1} + \beta_6^N \mathrm{lnKr}_{i,t-2} \\ &\quad + \beta_7^N \mathrm{lnZr}_{i,t-1} + \beta_8^N \mathrm{lnZr}_{i,t-2} + \beta_9^N \mathrm{lnD}_{i,t} + \beta_{10}^N \mathrm{lnD}_{i,t-1} + \varepsilon_{i,t}^N\end{aligned} \tag{9}$$

$$\begin{aligned}\mathrm{lnKr}_{i,t} &= \beta_1^K \mathrm{lnYr}_{i,t-1} + \beta_2^K \mathrm{lnYr}_{i,t-2} + \beta_3^K \mathrm{lnN}_{i,t-1} + \beta_4^K \mathrm{lnN}_{i,t-2} + \beta_5^K \mathrm{lnKr}_{i,t-1} + \beta_6^K \mathrm{lnKr}_{i,t-2} \\ &\quad + \beta_7^K \mathrm{lnZr}_{i,t-1} + \beta_8^K \mathrm{lnZr}_{i,t-2} + \beta_9^K \mathrm{lnD}_{i,t} + \beta_{10}^K \mathrm{lnD}_{i,t-1} + \varepsilon_{i,t}^K\end{aligned} \tag{10}$$

$$\begin{aligned}\mathrm{lnZr}_{i,t} &= \beta_1^Z \mathrm{lnYr}_{i,t-1} + \beta_2^Z \mathrm{lnYr}_{i,t-2} + \beta_3^Z \mathrm{lnN}_{i,t-1} + \beta_4^Z \mathrm{lnN}_{i,t-2} + \beta_5^Z \mathrm{lnKr}_{i,t-1} + \beta_6^Z \mathrm{lnKr}_{i,t-2} \\ &\quad + \beta_7^Z \mathrm{lnZr}_{i,t-1} + \beta_8^Z \mathrm{lnZr}_{i,t-2} + \beta_9^Z \mathrm{lnD}_{i,t} + \beta_{10}^Z \mathrm{lnD}_{i,t-1} + \varepsilon_{i,t}^Z\end{aligned} \tag{11}$$

$$\mathrm{lnD}_{i,t} = \rho_1 \mathrm{lnD}_{i,t-1} + \rho_2 \mathrm{lnD}_{i,t-2} + \varepsilon_{i,t}^D \tag{12}$$

We estimate these equations by OLS with fixed effects for each firm (FAD number) and we include time dummies to control for common unobserved macro shocks and trends.

As explained in Section 2 we think of the shocks in the first four equations as technology and cost shocks plus other shocks that we cannot measure. The key identifying

[12] More than 80% of the firms in the sample are in the same industry throughout the time that they exist in the data. A firm changing industry is assigned to the industry to which it belonged for the longest period of time. Typically, a firm does not change its production entirely but just passes a threshold in the composition of goods, leading to a change in industry classification.

assumption is that these shocks are uncorrelated with the demand variable after we have eliminated shocks that are common to all firms by including time dummies.

To see when this can be problematic, consider a simpler case where we have two types of firms producing investment and consumption goods. Then we identify demand shocks from the fact that an increase in investments raises demand only for firms producing investment goods. But suppose there is a technology or cost shock that affects a large fraction of the firms that produce investment goods.[13] Such a chock will reduce the price of investment goods and this will, most likely, lead to an increase in investments. Hence, this shock will affect firms that produce investment goods directly and also the demand variable as we measure it. To get some idea whether shocks of this type were important in this period we plot the ratio of the investment to consumption together with the ratio of the corresponding deflators in the Appendix. While we see a clear cyclical pattern in investment relative to consumption, the relative price varies much less and, if anything, the two variables are positively correlated. This means that, if anything, we underestimate the effects of demand shocks on production.

3.5 Impulse-response functions

Figure 2 shows what happens to the endogenous variables after an exogenous demand shock.[14] We see that the effect of the shock on demand is slightly hump-shaped and quite persistent; the half-time is 9 years. Since we have time dummies in the model, we are not capturing the general business cycle but rather shocks that are more persistent.

Production responds immediately to a change in demand and more than the demand shock itself, which may be because demand for goods produced in manufacturing is more volatile than aggregate consumption and investment (see discussion in Section 6). Production peaks one year after the peak of the shock. The inventory stock responds positively and with some lag: inventory holdings peak one year after the peak in production. The capital stock and employment respond with longer lags. The first-year effect on the number of workers is 23 percent of the effect on output and the capital stock does not respond at all in the first year. Thus we see that firms are able to satisfy demand with a relatively small increase in registered inputs and this implies very strong responses of labor and factor productivity as they are commonly measured. Employment reaches its peak 4 years after the peak in production while

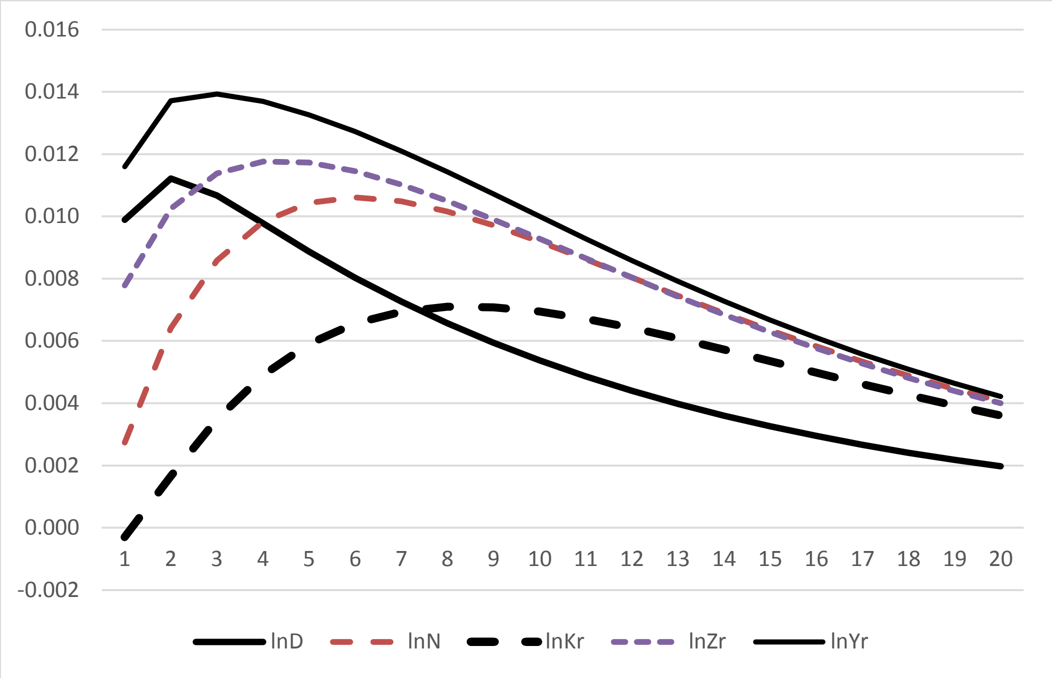
[13] For a macroeconomic analysis of the effects of investment-specific technology shocks, see Greenwood, Hercowitz and Krusell (2000).

[14] Confidence intervals for these impulse-response functions will be shown below. For a detailed presentation of the regression results, see the appendix.

the capital stock peaks 6 years after the peak in production. Note that employment almost catches up with production but the response of the capital stock is much weaker.

Figure 2. Impulse response functions for baseline sample

Balanced panel (used for estimation of theoretical model)



Note: Private manufacturing firms with at least ten employees and no extreme observations. Balanced panel: only firms with 13 observations. The variables are log deviations from the steady state, and the time units are years. Number of observations included in estimations: 8143.

3.6 Robustness

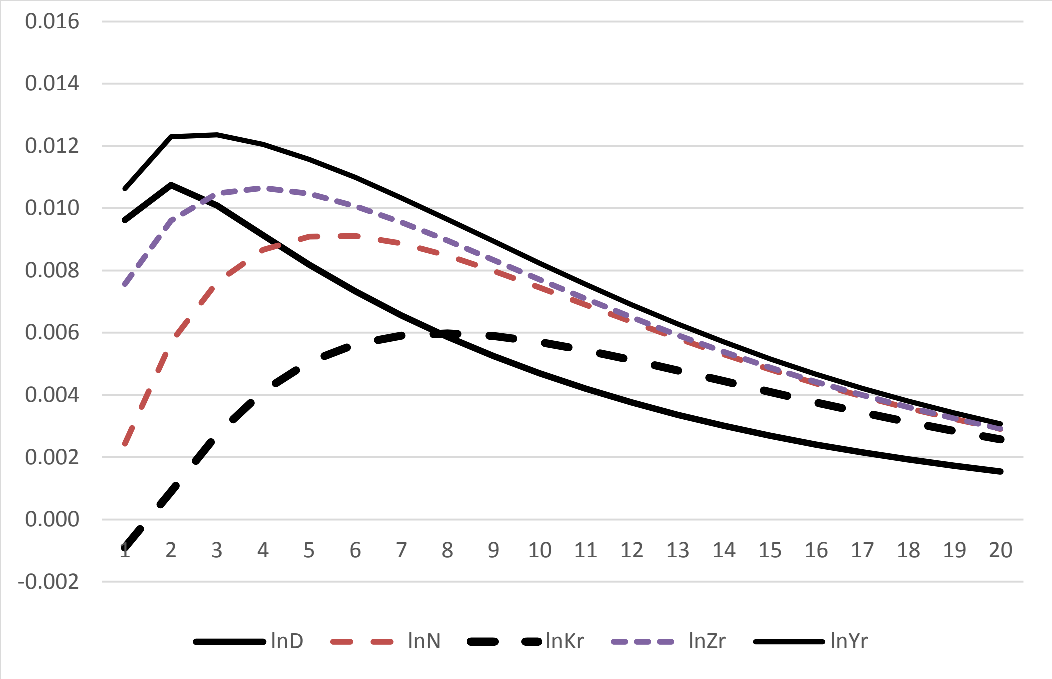
We now consider the results for alternative samples and specifications. *Figure 3a* shows the results for the unbalanced panel and *Figure 3b* shows the results for a sample where we exclude only extreme firm-year observations and not the whole time series for the firm with an extreme observation. In both cases, the results are very similar to our baseline results.

Figure 3c shows estimates where we use real value added as production measure instead of total production. Production and the capital stock respond more strongly in this case but otherwise the results are similar.

Figure 3d shows impulse-response functions from a model where we include 3 lags in the estimated empirical model. The responses of production and inventories become more hump-shaped, peaking 2 years after the peak in demand.

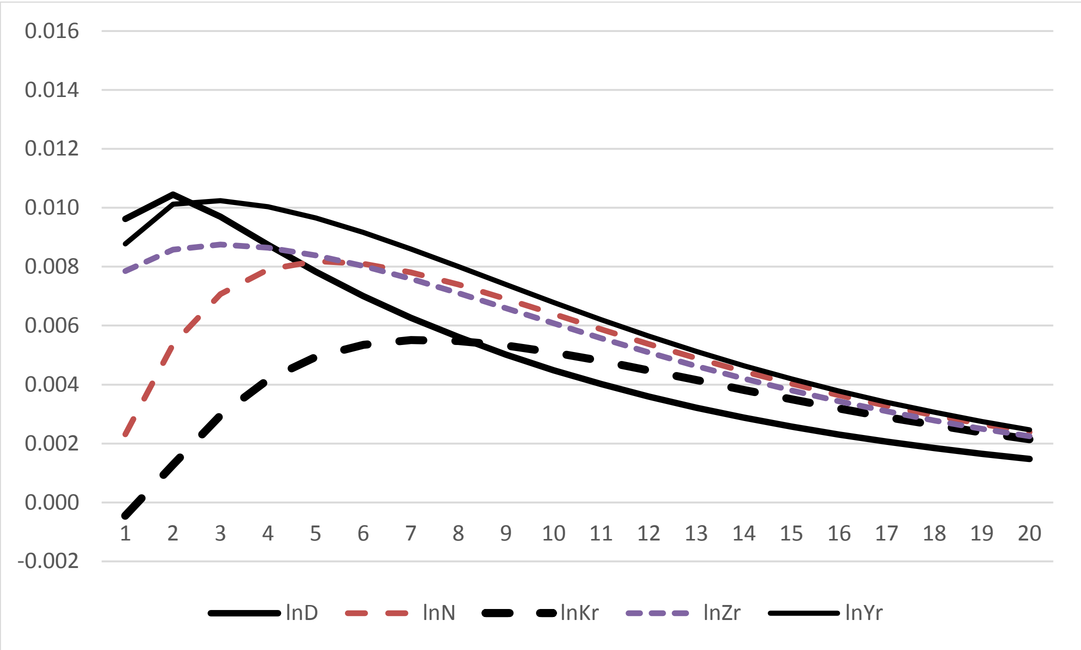
Figure 3. Impulse response functions: robustness

a) Unbalanced panel



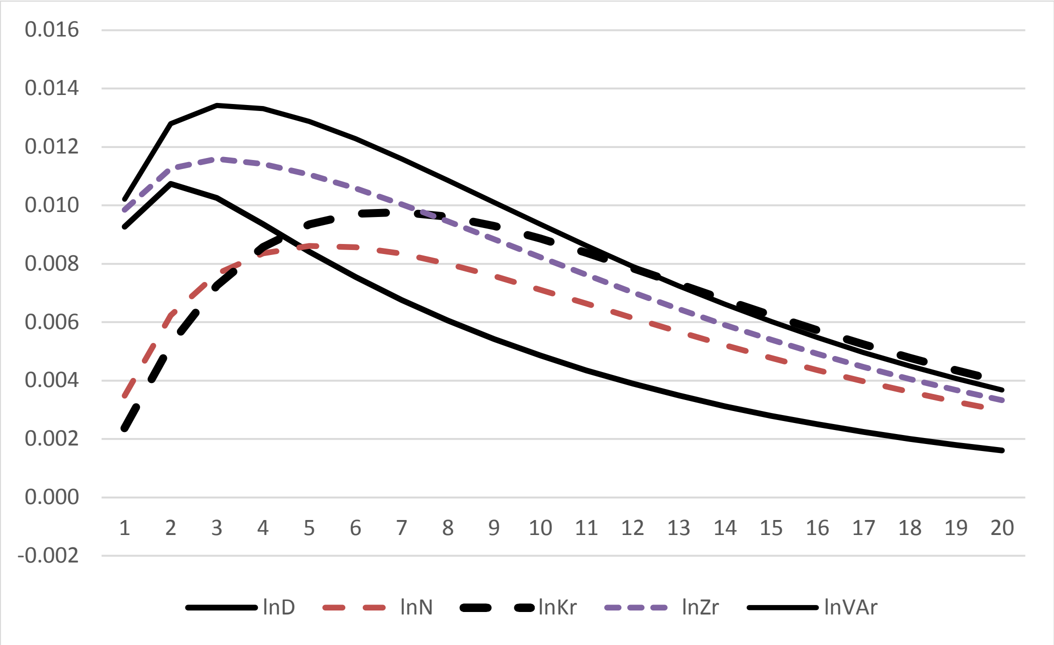
Note: Private manufacturing firms with at least ten employees and no extreme observations. Number of observations included in estimations: 11 306.

a) Leaving out extreme observations instead of firms with extreme observations



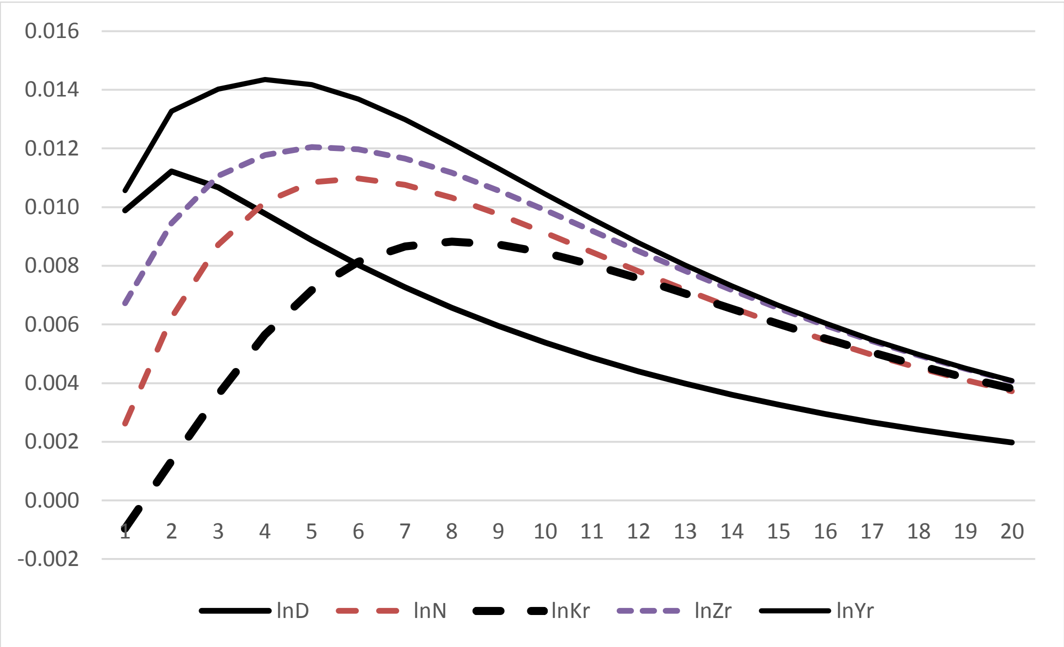
Note: Private manufacturing firms with at least ten employees. Extreme observations excluded but we use data for other years for firms with extreme observations. Number of observations used in estimations: 17 179.

b) Using value added as measure of production



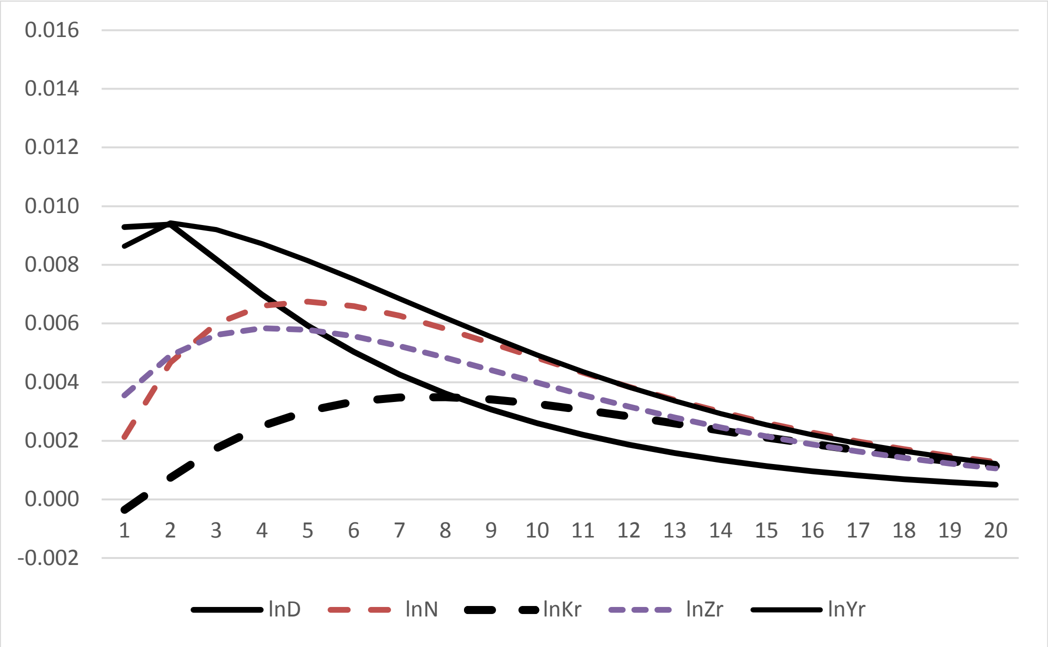
Note: . Balanced panel of private manufacturing firms with at least ten employees. Number of observations: 6547.

c) Including three lags of endogenous variables



Note: Balanced panel of private manufacturing firms with at least ten employees. Number of observations: 7325.

d) Including linear industry trends



Note: Balanced panel of private manufacturing firms with at least ten employees. Number of observations: 8143.

e) Including linear industry trends but no time dummies

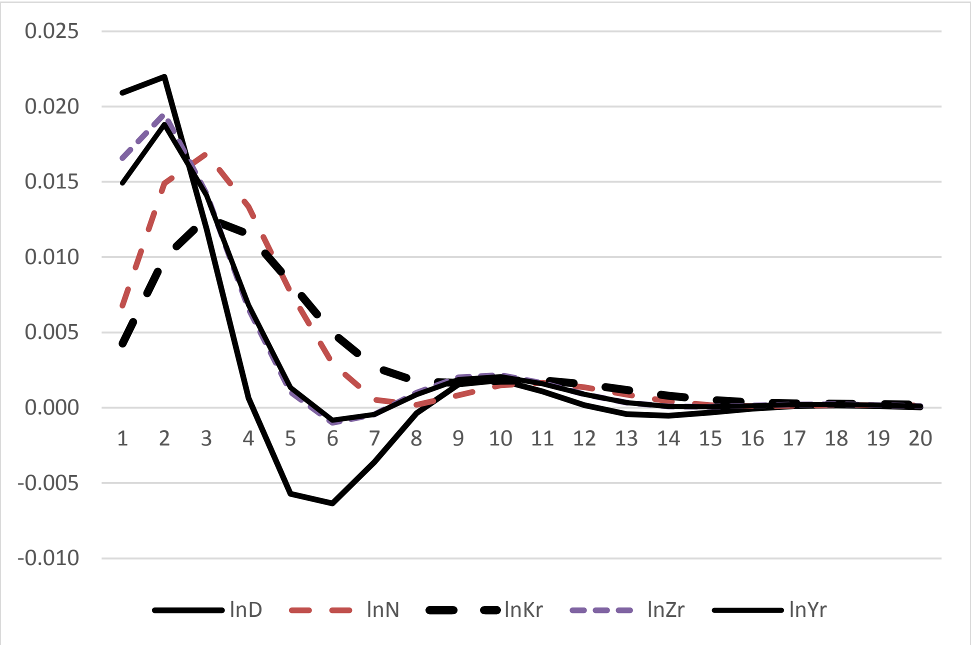


Figure 3e shows IRFs from a model where we have included industry-specific linear trends in the model. The overall picture is similar but the shock becomes less persistent when some of the industry variation is mopped up by the trend, the half-time falls to 7 years, and the inventory response becomes weaker.

Figure 3f shows the results when we include industry trends but exclude time dummies. Now, the shock is much less persistent, reflecting business cycle variation to a greater extent. As explained above, we do not use this variation for estimation because of the risk of spurious correlation due to unobserved aggregate shocks.

To sum up, we find that production, inventories, labor and the capital stock respond positively to the demand shock. While production responds quickly, labor and the capital stock respond slowly. The time profiles vary a bit for different specifications and the time profile of the inventory response is less robust compared to the other responses.

We also divided the sample into larger and smaller firms where larger firms are defined as having mean employment of at least 50. The resulting IRFs are reported in the Appendix, *Figure A2*. The responses of larger firms are similar to the baseline estimation but larger firms appear to be more cyclical in the sense that they respond more to the demand shock. Most likely, large firms produce goods that are more investment-type and durable goods. Smaller firms respond much less to the demand shock and the response of the capital stock is weaker. One reason may be that small firms often rent the capital that they use. Clearly, there is interesting heterogeneity among firms and our baseline estimation should be seen as a rough estimate of the average response across firms.

4. Theory

There appear to be some adjustment lags or costs that slow down the adjustment of capital and labor input but production and inventories respond quickly to the demand shock. Below we present some features of our theoretical model, which could potentially help to match these empirical responses. First, we discuss adjustment costs, implementation lags, increasing returns to scale, factor utilization and price rigidity. Then we specify the relation between output and value added and our model of inventory holdings. The firm's maximization problem is presented in section 4.7.

4.1 Adjustment costs and implementation lags

We include standard *quadratic adjustment costs* for labor and capital. The adjustment costs are equal to $c_N\left(H_t - \delta_n \bar{N}\right)^2/2 + c_K\left(I_t - \delta_k \bar{K}\right)^2/2$ where δ_k is the rate at which capital depreciates and δ_n is an exogenous separation rate for labor. $\bar{N}$ and $\bar{K}$ denote the steady state levels of N_t and K_t so there are quadratic costs for hiring and investing more than the steady state levels $\delta_n\bar{N}$ and $\delta_k\bar{K}$.[15] These costs take the form of reduced production due to disruptions in the production process.

We also include *implementation lags in investments and hiring* by assuming that some given fraction of the investments and hiring that are *decided* in year t are *implemented* in year t+1:

$$K_t = \left(1-\delta_k\right)K_{t-1} + \lambda_k I_t + \left(1-\lambda_k\right)I_{t-1} \tag{13}$$

$$N_t = \left(1-\delta_n\right)N_{t-1} + \lambda_n H_t + \left(1-\lambda_n\right)H_{t-1}. \tag{14}$$

K_t is the capital stock at the end of the period and I_t is investments *decided* in period t. N_t is employment during period t and H_t is hiring *decided* in period t. The idea is that if you decide to make a specific investment in a particular year, some of that investment will be implemented in the current year and some will be implemented in the next year and the same applies to hiring. This is similar to "time to build" (Kydland and Prescott 1982) and more flexible than assuming either no lag or a one-period implementation lag as in Burnside-Eichenbaum-Rebelo (1993).

4.2 Increasing returns to scale in production

Our estimated impulse-response functions show that firms can increase production in the short run with much smaller percent increases in registered inputs of capital and labor. The theoretical model includes several features that can help to explain this. One possible explanation, is that there are *increasing returns to scale* so that changes in inputs lead to proportionally larger changes in production (Hall 1988). To model this, we assume that the capital stock consists of two components. First, there is a flexible part $\hat{K}_t$ which enters a CES production function with constant returns to scale and second there is a fixed amount of capital F_k that the firm must have in order to produce at all. Thus total observed capital is

[15] This specification is chosen to make it simpler to solve analytically for the steady state but it should not affect the dynamics because a positive marginal adjustment cost for labor in steady state is equivalent to a higher wage.

given by $K_t = F_k + \hat{K}_t$. Similarly, we distinguish between fixed and flexible employment: $N_t = F_n + \hat{N}_t$.

4.3 Factor utilization and organizational capital

Looking at the dynamic response in *Figure 2*, we see that production increases much more than observed inputs in the first year, but after some years, employment has almost caught up with production. It is unlikely that increasing returns can explain the whole picture; firms appear to have some form of excess capacity that they can use to meet demand. A standard way to allow for this is to allow for variable *utilization* of the factors of production (see e.g. Burnside-Eichenbaum-Rebelo, 1993).

The key question, then, is why the firm did not make full use of its resources for production before the shock occurred. Clearly, there must be some cost of increasing resource utilization or else the firm would always make full use of its resources in production. Some researchers have modelled variations in the utilization of capital assuming that increased use of capital makes the capital stock depreciate at a faster rate when it is heavily utilized. Others have allowed for variations in workers' effort where the firm has to compensate the workers for increasing their effort in times of high demand or productivity.

We allow for variations in utilization (u_t) of both factors of production at a cost given by $\Phi_u \left(\left(1 - c_u\right) u_t + c_u \cdot u^2 / 2 \right) \hat{N}_t$. The variable u_t enters multiplicatively in the production function below; it may represent effort or overtime, which increases the use of both labor and capital. In the latter interpretation, the convex cost may reflect an overtime premium, which may be part of an explicit or implicit contract.[16]

Several authors have noted that workers spend substantial amounts of time on activities, which do not contribute to *current* production, but which increase *future* production (e.g. Fay and Medoff (1985)). There are many such activities we can think of: cleaning and maintenance, reorganizing, training etc. To capture this, we include another element in the model that we call *organizational capital*. We assume that the firm has a stock of organizational capital Ω and the larger this stock is, the more it can produce with given inputs. Workers spend a share x_t of their time on activities that increase current production

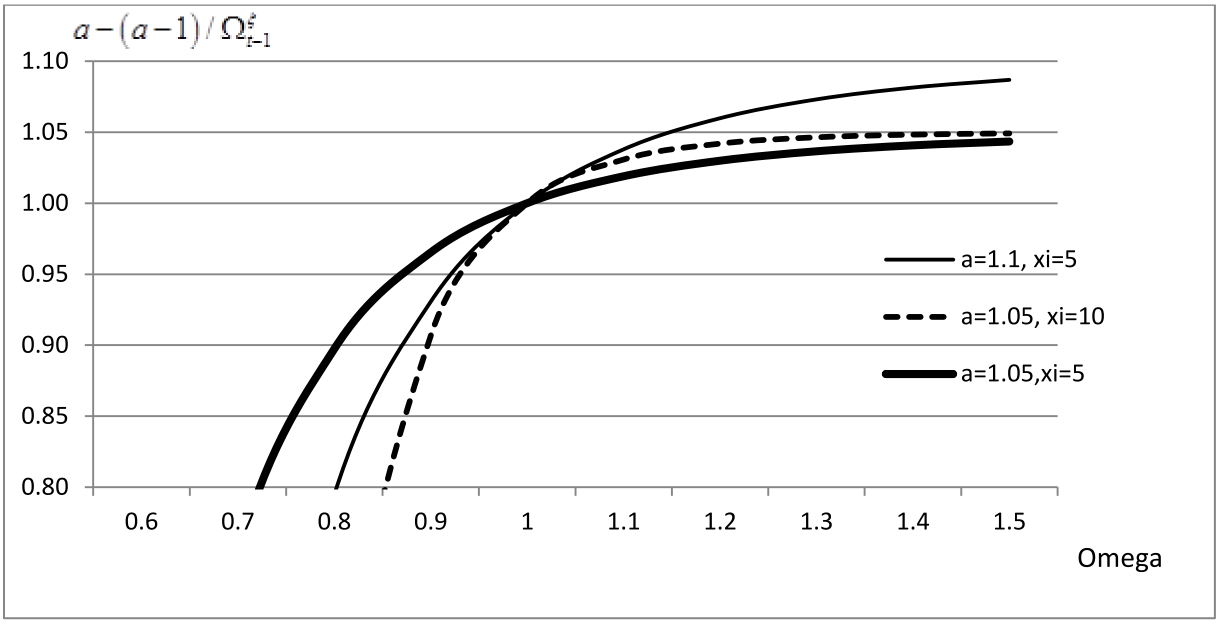
[16] As mentioned in Section 3, the number of workers is measured as "full-time equivalents" and we suspect that this measure does not fully account for variations in hours, so variations in u_t will partly reflect variations in hours per worker.

and a fraction $1-x_t$ of their time accumulating organizational capital, which increases future production. Thus, we write the production function for value added

$$F\left(A_t, u_t, \Omega_{t-1}, \hat{K}_{t-1}, x_t, \hat{N}_t, H_t, I_t\right) = A_t u_t \left(a - \frac{a-1}{\Omega_{t-1}^{\xi}}\right)\left(\alpha \hat{K}_{t-1}^{\frac{\sigma-1}{\sigma}} + (1-\alpha)\left(x_t \hat{N}_t\right)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} - \frac{c_N}{2}\left(H_t - \delta_n \bar{N}\right)^2 - \frac{c_K}{2}\left(I_t - \delta_k \bar{K}\right)^2 \quad (15)$$

where $a>1$ and where $\hat{K}_t$ and Ω_t denote the stocks of physical and organizational at the end of period t. *Figure 4* illustrates the function $a-(a-1)/\Omega_{t-1}^{\xi}$ for different values of the parameters *a* and ξ. We normalize so that $\Omega=1$ in steady state and thus the function value is one in steady state. As organizational capital increases, the function value increases asymptotically towards a and if organizational capital falls to $(a-1)/a$ the function value falls to zero. Roughly speaking, the parameter *a* determines the slope of the function while ξ determines its concavity.

Figure 4. Productivity contribution of organizational capital



Note: The function shows normalized factor productivity (y-axis) as a function of organizational capital (Ω) for different values of the parameters a and ξ.

We assume that the accumulation of organizational capital is governed by

$$\Omega_t = (1-\delta_\omega)\Omega_{t-1} + \chi(1-x_t) \tag{16}$$

where δ_ω is the depreciation rate of organizational capital $1-x_t$ is the fraction of time spent accumulating of organizational capital. The parameter χ is set to be consistent with the normalization that $\Omega = 1$ in steady state.

The basic idea behind this specification is that, when there is temporarily high demand, the firm may tell the workers to increase the fraction of their time spent on current production and to spend less time on maintenance and training. Note that variations in effective work hours (u_t) and variations in time spent investing in organizational capital (x_t) have similar effects on production today. The difference is that variations in utilization are associated with a direct cost today, while disinvestment in organizational capital shows up as lower productivity in the future.

4.4 Output and value added

We assume that value added and materials inputs are combined in a Leontief production function

$$Y_t = \min\left\{F\left(A_t, u_t, \Omega_{t-1}, \hat{K}_{t-1}, x_t, \hat{N}_t, H_t, I_t\right), M_t/\mathrm{m}\right\}$$

where Y_t is the quantity produced, $F(\cdot)$ was defined above and M_t is the quantity of inputs. Cost minimization then implies that $Y_t = F(\cdot) = M_t/m$. Normalizing the price of inputs to one, the cost of inputs is mY_t; you need a fixed amount of cloth to make a shirt.

Since there is no substitutability between value added and materials, materials inputs and total output will always be proportional to value added and it should not matter whether we measure production by output or value added. Clearly, one could allow for some substitutability between materials and other inputs, but as seen in *Figure 3,* the impulse-response functions are fairly similar when we use value added to measure production. Note that we allow for increasing returns by including fixed costs in terms of capital and labor, but not in terms of materials.[17]

[17] See section VI in Basu (1996) and Basu-Fernald (1997) page 255, for discussions of these issues.

4.5 Price rigidity

Another factor that can prevent firms from always optimally utilizing their resources is *price rigidity*. If demand falls and the firm cannot (or does not want to) reduce its price, quasi-fixed resources will become less utilized (Rotemberg and Summers 1990). We incorporate price rigidity in a simple way by including a quadratic adjustment cost for prices and we assume that firms always satisfy demand, which makes sense if firms have sufficient market power.

4.6 Inventory model

The estimated responses show that firms increase their inventory stocks when demand increases. This is opposite to the production-smoothing idea that, by drawing down inventories in periods of high demand, firms can make production more stable. To explain the observed pattern, we follow Kahn (1987, 1992) and Bils and Kahn (2000) and think of inventories more like a productive factor: inventories of finished goods are needed in order to sell the good. Below we present a very stylized model that gives us a reasonable functional form that we can include in our estimated model.[18]

Consider a firm that sells goods, e.g. shirts, which come in *M* different varieties, which we will call sizes. Let us assume that a customer will only buy the good if he/she finds the right size. The firm has a sales department and a production department and the sales department sends an order to the production department T times per year so as to replenish the inventory stock of finished goods. To fix ideas in the presentation below, we can think of the case when *T*=12 so inventories are replenished every month.[19]

Let $\hat{D}$ be *potential sales* of all varieties during a year (we omit time index here). $\hat{D}$ is what the firm would sell if it would never stock out and we assume that $\hat{D} = \Phi D^{\Sigma} P^{-\eta}$ where Φ is a constant, *D* is a demand shifter and *P* is the price set by the firm. To make the model as simple as possible, we assume that D is known, that *P* is set at the beginning of the year and that both are constant over the year. Demand in a particular month for a particular size is assumed to be $\lambda \hat{D}$ where λ is a stochastic variable that is uniformly distributed between λ_1 and λ_2. The supports of the distribution are given by $\lambda_1 = (1-\Psi)/(TM)$ and $\lambda_2 = (1+\Psi)/(TM)$ where *T* is the number of inventory periods (months) and *M* is the

[18] The first version of this inventory model was developed in Stadin (2014).
[19] Contrary to the Ss model, *T* is taken as exogenous, so our model differs from the Ss model in the same way as the Taylor model of wage/price setting differs from state-contingent pricing.

number of varieties. The parameter Ψ has a value between zero and unity and it reflects the degree of uncertainty about demand for individual varieties.

Since demand is symmetrically and independently distributed across sizes, the sales department will make sure they stock up with the same quantity of each size whenever they replenish inventories. Let *z* be the inventory stock of a specific size held at the beginning of a month. It follows immediately that $\lambda_1\hat{D} < z < \lambda_2\hat{D}$. With a smaller inventory stock you would always stock out and there is no reason to hold a larger inventory stock than maximum possible sales of a particular size. If the realization of λ is such that $\lambda\hat{D} \leq z$, sales of that specific size will be $\lambda\hat{D}$, and if $\lambda\hat{D} > z$, sales of that specific size will be *z*. Letting $\hat{\lambda}$ be the critical value of λ where the firm runs out of stock $\left(\hat{\lambda}\hat{D} = z\right)$ we get expected sales of a particular size in a given month as

$$s = \int_{\lambda_1}^{\hat{\lambda}} \lambda\hat{D}\frac{TM}{2\Psi}d\lambda + \int_{\hat{\lambda}}^{\lambda_2} z\frac{TM}{2\Psi}d\lambda = \frac{TM}{2\Psi}\left(\frac{\hat{\lambda}^2\hat{D}}{2} - \frac{\lambda_1^2\hat{D}}{2} + \lambda_2 z - \hat{\lambda}z\right) = \frac{1+\Psi}{2\Psi}z - \frac{\left(1-\Psi\right)^2}{4\Psi TM}\hat{D} - \frac{TM}{4\Psi}\frac{z^2}{\hat{D}} \tag{17}$$

We assume that $\hat{D}$ is observed at the beginning of the year and constant over the year, so total expected sales of storable finished goods over the year are:

$$TMs = \frac{1+\Psi}{2\Psi}T\hat{Z} - \frac{\left(1-\Psi\right)^2}{4\Psi}\hat{D} - \frac{T^2\hat{Z}^2}{4\Psi\hat{D}} \tag{18}$$

where $\hat{Z}$ is the total stock of finished goods at the beginning of the month: $\hat{Z} = Mz$. To keep the model simple, we assume that the firm has a large number of varieties so we can view this function as a deterministic function that determines sales. This function has several natural properties:

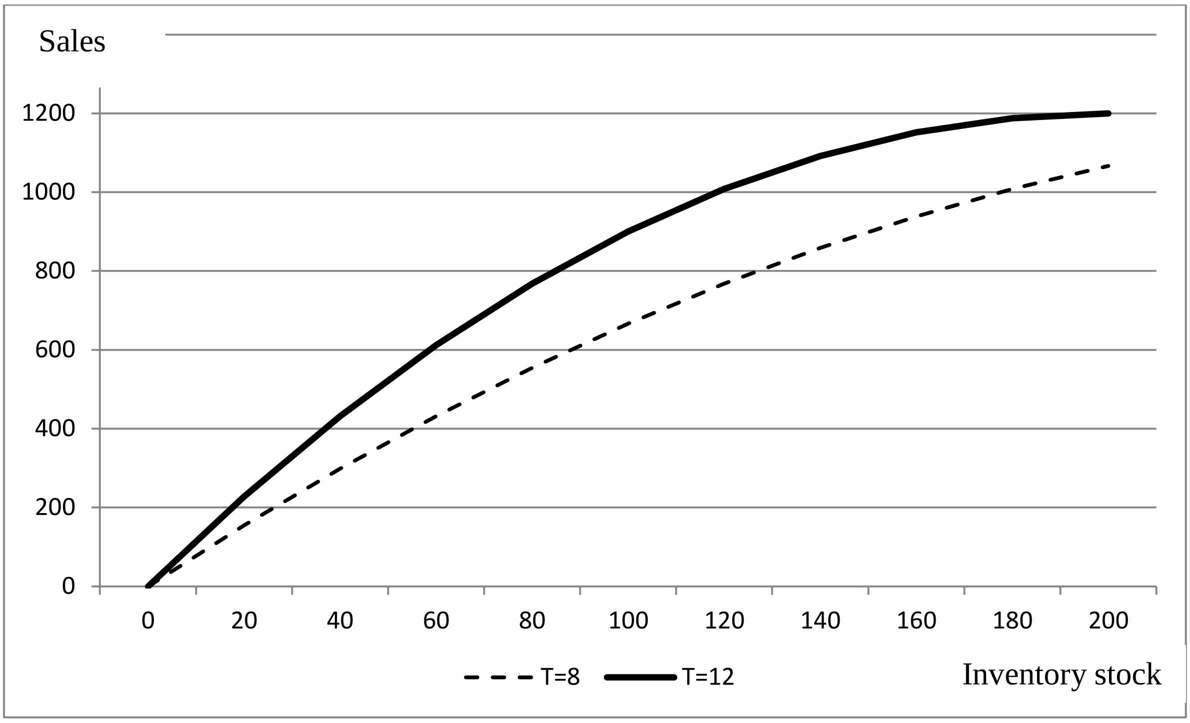
i) For a given inventory stock, maximum sales are $T\hat{Z}$ and sales approach that limit as $\hat{D}$ goes to $\hat{Z}T/\left(1-\Psi\right)$; for a lower level of demand there will be some varieties which will not sell out.

ii) For a given level of demand, the inventory stock that maximizes sales is $\hat{Z} = \left(1+\Phi\right)\hat{D}/T$: in order to make sure that all the required sizes are always available you need inventories of each variety corresponding to maximum demand during an inventory period. (In the model, costs of financing, depreciation and storage make optimal inventories smaller than this.)

iii) Starting from a low T, a higher T will increase sales as you replenish the inventory stock more often.

Figure 5 shows sales of finished storable goods (TMs) as a function of the inventory stock setting potential sales to 1200, T=12 so the stock is replenished every month, and $\Psi = 1$ so that demand for a particular size is uniformly distributed between zero and twice expected demand for that size. In this case, sales are equal to $12 \cdot \hat{Z} - 0.03 \cdot \hat{Z}^2$ in the relevant interval. In order to never stock out the firm needs to have an inventory stock which is 200, i. e. twice as large as potential sales. With a smaller stock it will sell less because some sizes will run out. If the stock is replenished more seldom this will reduce sales for a given stock.

Figure 5. Sales of storable goods as a function of the inventory stock $\left(\Psi = 1, \hat{D} = 100\right)$



Note: T denotes the number of times you replenish the inventories per year.

To this we add yet another modification by assuming that there are some goods that are sold without holding stock. These may be perishable goods or goods produced on order. Sales of these are simply assumed to be equal to $\hat{D}$. Letting the fraction of storable finished goods be Λ, we get total sales as

$$S = \Lambda TMs + \left(1 - \Lambda\right)\hat{D} = \kappa_1 \hat{Z} + \kappa_2 D^{\Sigma} P^{-\eta} - \kappa_3 \hat{Z}^2 D^{-\Sigma} P^{\eta} \tag{19}$$

where $\kappa_1 = \Lambda(1+\Psi)\mathrm{T}/(2\Psi)$, $\kappa_2 = 1-\Lambda-\Lambda(1-\Psi)^2\Phi/(4\Psi)$ and $\kappa_3 = \Lambda T^2/(4\Psi\Phi)$.

The accumulation of finished goods inventories is governed by the function

$$\hat{Z}_t = (1-\delta_z)\hat{Z}_{t-1} + Y_t + S_t \tag{20}$$

where $\hat{Z}_t$ is the finished goods inventory stock at the end of the year and δ_z is the rate at which inventories depreciate during the year. We also include a cost $c_z \cdot \hat{Z}_t$ which reflects other costs of holding inventories such as the cost of providing storage space and managing inventories.

Finally, we note that inventories consist not only of finished goods, but also of inputs and goods in process and our data do not allow us to distinguish different types of inventories. To take account of this we simply assume that the firm holds a stock of inputs which is proportional to current production: $h_z Y_t$. These inputs can be bought without delay and hence the total observed inventory stock, Z_t is given by $Z_t = \hat{Z}_t + h_z Y_t$.

4.7 Profit maximization

The firm's profit maximization problem is to choose $S_t, Y_t, P_t, \hat{K}_t, I_t, \hat{N}_t, H_t, u_t, x_t, \Omega_t, \hat{Z}_t$ to maximize

$$E_t\left\{\sum_{\tau=t}^{\infty}\beta^{\tau-t}\left[\begin{array}{l} S_\tau P_\tau - W_\tau \hat{N}_\tau - mY_t - \Phi_u\left(u_t - 1 + \frac{c_u}{2}(u_t-1)^2\right)\hat{N}_\tau \\ -P_\tau^K\left(\hat{K}_\tau - (1-\delta_k)\hat{K}_{\tau-1}\right) - c_z\hat{Z}_t - \frac{\theta}{2}\left(\frac{P_\tau}{P_{\tau-1}}-1\right)^2 \end{array}\right]\right\} \tag{21}$$

subject to the following constraints (with associated shadow prices) which hold for all t:

$$Y_t = A_t u_t\left(a - \frac{a-1}{\Omega_{t-1}^{\xi}}\right)\left(\alpha\hat{K}_{t-1}^{\frac{\sigma-1}{\sigma}} + (1-\alpha)\left(x_t\hat{N}_t\right)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} - \frac{c_N}{2}\left(H_t - \delta_n\bar{N}\right)^2 - \frac{c_K}{2}\left(I_t - \delta_k\bar{K}\right)^2 \qquad \nu_t \tag{22}$$

$$S_t = \kappa_1\hat{Z}_t + \kappa_2 D_t^{\Sigma}P_t^{-\eta} - \kappa_3\hat{Z}_t^2 D_t^{-\Sigma}P_t^{\eta} \qquad \mu_t \tag{23}$$

$$\hat{Z}_t = (1-\delta_z)\hat{Z}_{t-1} + Y_t - S_t \qquad g_t \tag{24}$$

$$F_k + \hat{K}_t = (1-\delta_k)\left(F_k + \hat{K}_{t-1}\right) + \lambda_k I_t + (1-\lambda_k)I_{t-1} \qquad q_t \tag{25}$$

$$F_n + \hat{N}_t = (1-\delta_n)\left(F_n + \hat{N}_{t-1}\right) + \lambda_n H_t + (1-\lambda_n)H_{t-1} \qquad \gamma_t \tag{26}$$

$$\Omega_t = (1-\delta_\omega)\Omega_{t-1} + \chi(1-x_t). \qquad \phi_t \tag{27}$$

Defining $\hat{Y}_t = Au_t\left(a-\frac{a-1}{\Omega_{t-1}^{\xi}}\right)\left(\alpha\hat{K}_{t-1}^{\frac{\sigma-1}{\sigma}}+(1-\alpha)\left(x_t\hat{N}_t\right)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$ we can write the first order conditions:

$$S_t:\quad \mu_t = P_t - g_t \tag{28}$$

$$Y_t:\quad g_t = v_t + m \tag{29}$$

$$P_t:\quad S_t - \mu_t\eta\left(\kappa_2 D_t^{\Sigma}P_t^{-\eta-1}+\kappa_3\hat{Z}_t^2 D_t^{-\Sigma}P_t^{\eta-1}\right)-\theta\left(\frac{P_t}{P_{t-1}}-1\right)\frac{1}{P_{t-1}}+\beta\theta E_t\left\{\left(\frac{P_{t+1}}{P_t}-1\right)\frac{P_{t+1}}{P_t^2}\right\}=0 \tag{30}$$

$$K_t:\quad q_t = -P_t^K + \beta E_t\left\{v_{t+1}\alpha\left(\hat{Y}_{t+1}/\hat{K}_t\right)^{\frac{1}{\sigma}}+(1-\delta_K)\left(P_{t+1}^K+q_{t+1}\right)\right\}$$

(31)

$$I_t:\quad v_t c_k\left(I_t-\delta_k\bar{K}\right)=\lambda_k q_t+\beta(1-\lambda_k)E_t\left(q_{t+1}\right) \tag{32}$$

$$\hat{N}_t:\quad \gamma_t = v_t(1-\alpha)x_t^{\frac{\sigma-1}{\sigma}}\left(\hat{Y}_t/\hat{\mathrm{N}}_t\right)^{\frac{1}{\sigma}}-W_t-\Phi\left(u_t-1+\frac{c_u}{2}\left(u_t-1\right)^2\right)+\beta(1-\delta_n)E_t\left(\gamma_{t+1}\right) \tag{33}$$

$$H_t:\quad v_t c_n\left(H_t-\delta_n\bar{N}\right)=\lambda_n\gamma_t+\beta(1-\lambda_n)E_t\left(\gamma_{t+1}\right) \tag{34}$$

$$u_t:\quad \Phi_u\left(1-c_u+c_u u_t\right)\hat{N}_t = v_t\hat{Y}_t/\mathrm{u}_t \tag{35}$$

$$x_t:\quad \phi_t\chi = v_t(1-\alpha)\hat{N}_t^{\frac{\sigma-1}{\sigma}}\left(\hat{Y}_t/x_t\right)^{\frac{1}{\sigma}} \tag{36}$$

$$Z_t:\quad g_t=\mu_t\left(\kappa_1-2\kappa_3\frac{\hat{Z}_t}{D_t^{\Sigma}P_t^{-\eta}}\right)-c_z+\beta(1-\delta_z)E_t\left(g_{t+1}\right) \tag{37}$$

$$\Omega_t:\quad \phi_t=\beta E_t\left\{v_{t+1}\frac{\hat{Y}_{t+1}(a-1)\xi}{\left(a-(a-1)\Omega_t^{-\xi}\right)\Omega_t^{\xi+1}}+(1-\delta_{\omega})\phi_{t+1}\right\}. \tag{38}$$

The total amounts of capital, labor and inventory stock are $K_t = F_k + \hat{K}_t$, $N_t = F_n + \hat{N}_t$ and $Z_t = h_z Y_t + \hat{Z}_t$. Finally, we also need to specify the stochastic process for the demand shock. In line with our empirical model, we assume that the demand shock follows an AR(2) process

$$D_t = 1+\rho_1\left(D_{t-1}-1\right)+\rho_2\left(D_{t-2}-1\right)+\varepsilon_t \tag{39}$$

To sum up, adjustment costs and implementation lags may help to explain sluggish adjustment of labor and capital while increasing returns to scale in production, variable utilization and (dis)investments in organizational capital could potentially explain the observed increase in factor productivity (as it is normally measured) in response to a demand shock. A positive inventory response may arise because of the stock-out motive and because a large fraction of inventories consist of inputs and goods in process. We now turn to the estimation which will help us to discriminate between these alternative explanations.

5. Estimation method

We follow Christiano, Eichenbaum and Evans (2005) and estimate the structural parameters in the theoretical model by finding the set of parameter values that makes the impulse responses in the theoretical model match the impulse responses of its empirical counterpart.

5.1 Matching impulse response functions

The target function is constructed as in Christiano, Eichenbaum and Evans (2005):

$$J = \min\left[\hat{\Psi} - \Psi\left(\vec{\gamma}\right)\right]V^{-1}\left[\hat{\Psi} - \Psi\left(\vec{\gamma}\right)\right]. \tag{40}$$

$\Psi\left(\vec{\gamma}\right)$ contains the impulse responses calculated with the theoretical model for different horizons as a function of the model parameter vector $\vec{\gamma}$ and $\hat{\Psi}$ is the empirical counterpart. *V* is a diagonal matrix with the variances from the empirical estimation. These variances are related to the 95% confidence intervals which are shown in *Figure 6* below. We include 20 years of IRFs in the estimation.

5.2 Prior constraints

We constrain the parameters to be in an economically reasonable range. The columns denoted min and max in *Table 2* below show the prior intervals for the estimated parameters. Also, we impose some restrictions on the steady state levels to make sure that these are roughly consistent with what we know about the levels from accounting data.[20] In the balanced panel sample, the median cost of personnel relative to value added is 78 percent and we constrain this ratio to be between 73 and 83 percent in the steady-state. The median real capital (calculated as described above) relative to value added is 84 percent and we constrain this ratio to be between 79 and 89 percent in the steady state.

The median ratio of inventories to production is 15 percent but we should note that $\hat{Z}_t$ in the model is the stock of finished goods when the firm has just filled up the inventory stock. This means that the steady state level of inventories should be higher than the number

[20] A set of parameter estimates, which is grossly inconsistent with what we know about the levels is uninteresting.

observed in the data. We therefore constrain Z_t / Y_t to be between 15 and 25 percent in the steady state. We cannot distinguish different types of inventories in our data but we know the proportions of finished goods, inputs and goods in process for manufacturing as a whole. If we count goods in process as half finished goods and half inputs, then roughly half the inventory stock consists of finished goods. Therefore we constrain the ratio of finished goods to stored inputs to be between 2/3 and 3/2. Finally, we check that profits are positive in steady state.

5.3 Search algorithm

We use the search algorithm from Mickelsson (2016) which is based on the local algorithm of Nelder and Mead (1965). The basic idea of this algorithm is to start with a large number of starting vectors which are spread out across the parameter space and then to combine these vectors in a smart way so as to approach the global maximum without getting stuck at local maxima or iterating too long on flat surfaces. Mickelsson (2016) shows that this algorithm does better than most commonly used search algorithms when the objective function is tricky.

5.4 Confidence intervals

In order to get confidence intervals for the parameters, we generated distributions of the estimates in the following way:

1. First we create a new sample of firms of the same size as the original sample by drawing firms randomly from the original sample with replacement.
2. This sample is used to get a new estimate of the empirical IRFs.
3. The impulse responses from the empirical model are then used to estimate the parameters of the theoretical model, as described above.
4. The vector of parameter estimates is saved and steps 1-3 are repeated 1000 times to get a distribution of estiamtes.[21]

6. Results

6.1 Some parameters are poorly identified or have corner solutions

There are many parameters in our theoretical model so it is not surprising that some parameters are poorly identified from these data. Attempts to estimate the discount factor, the

[21] We tested bootstrapping with the restriction that the number of firms from each sector should be the same in each sample but this made very little difference for the results.

depreciation rates, price rigidity and the elasticity of substitution indicate that these parameters are hard to identify. For this reason, we set the discount rate (β) to 0.96, all depreciation rates $(\delta_k, \delta_n, \delta_z)$ to 0.11 and we assumed flexible prices $(\theta = 0)$. The elasticity of substitution (σ) was set to 0.4 based on estimates by Chirinko, Fazzari and Meyer (2011). Later, we show that the dynamic responses are not very sensitive to changes in these parameters, which explains why they are poorly identified.

Our preliminary estimates suggested that variations in organizational capital play a small role and we therefore decided to simplify the model by omitting this aspect. Thus we set $\Omega = x = 1$ and omitted the equations that relate to Ω and *x*. We discuss this result below.

6.2 Replication of empirical responses

Figure 6 shows that the model is able to replicate the empirical impulse response functions almost perfectly. Also shown are the confidence intervals for the empirical IRFs which have been calculated by bootstrapping (resampling the firms 1000 times with replacement).

Figure 6. Theoretical and empirical impulse response functions

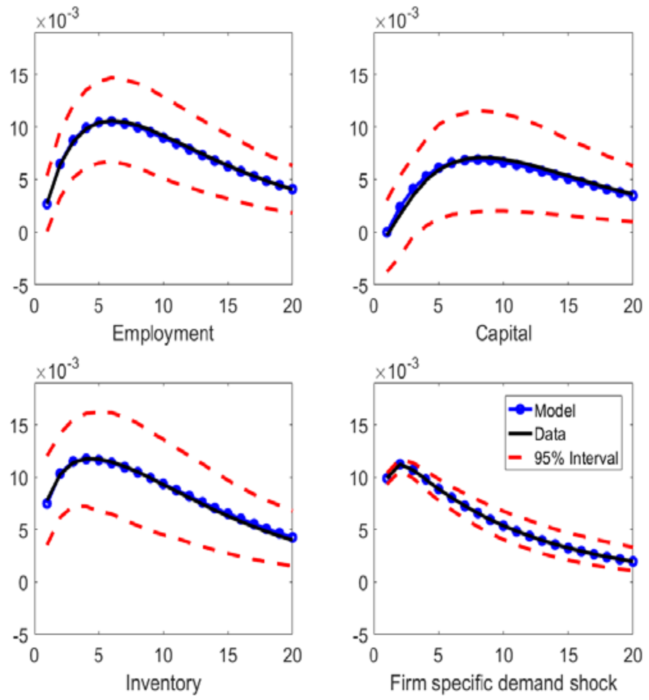
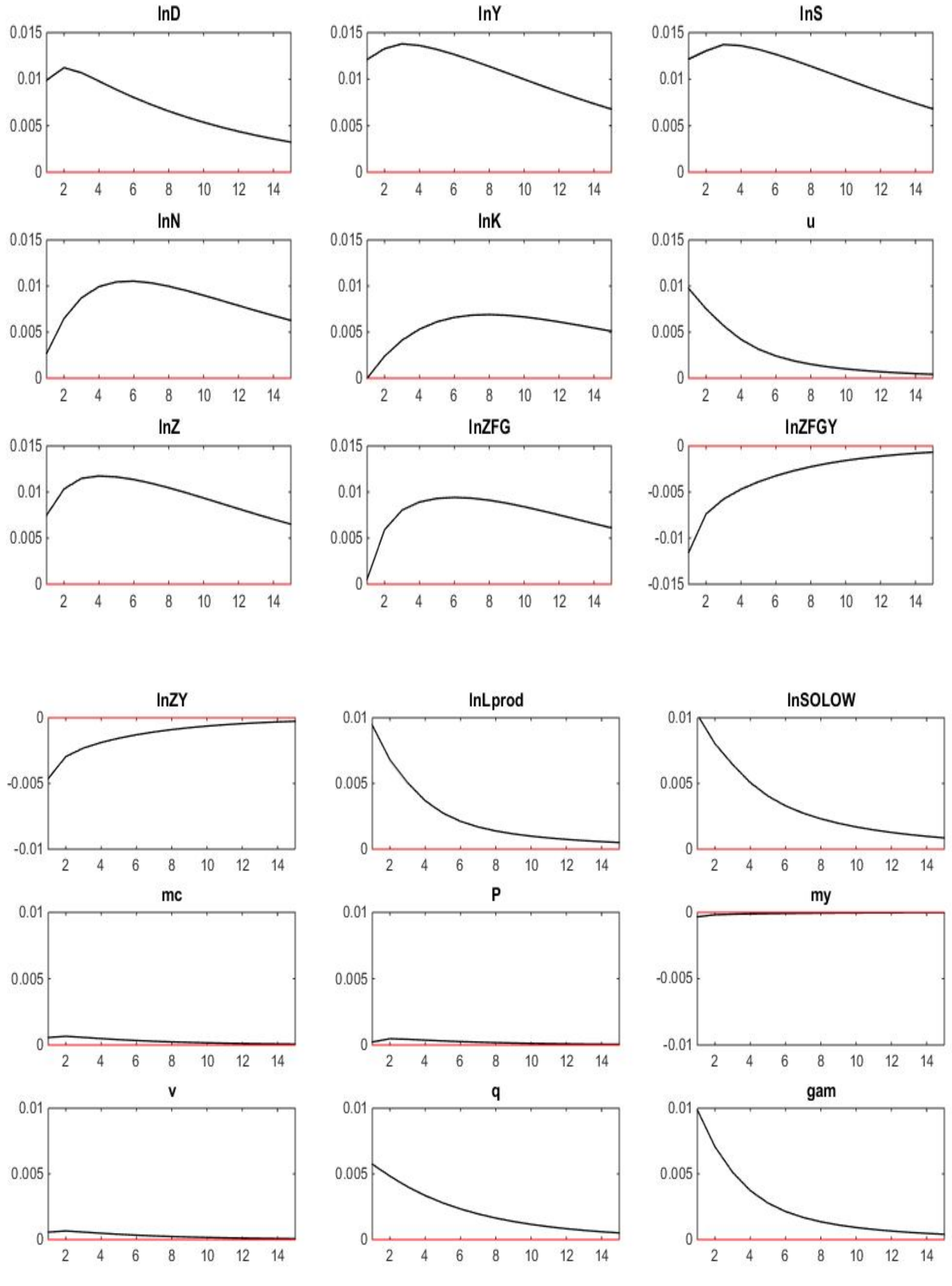


Figure 7. Variables in the theoretical model



Note : lnZFG=log of inventories of finished goods, lnZFGY=log of inventories of finished goods relative to production, $\ln\left(\hat{Z}\right)$ lnLprod=lnN-lnY, lnSOLOW=lnY-0.3*lnK-0.7*lnN,

Figure 7 shows the dynamics of some key theoretical variables for which we do not have data. As demand increases, production increases along with the demand shock. The initial increase in production is about 1.2 percent and most of this increase in production is achieved by increasing utilization by 1 percent while the rest is achieved by increasing employment by about 0.2 percent. Note that, since it is $\hat{K}_{t-1}$ that enters the production function, the capital stock does not contribute to the increase in production in the first or the second year. Slow adjustment of labor and capital is explained by substantial adjustment costs and implementation lags in hiring and investment. Labor and factor productivity respond strongly to the demand shock. According to the estimates, the cost of utilization is not very convex. This can be seen from the fact that marginal cost does not rise very much although labor and capital respond very sluggishly. Note also that the price increases even less than marginal cost so the markup declines in response to an increase in demand. The roles of the different parameters are discussed further below.

6.3 Parameter estimates and confidence intervals

One way to illustrate the precision of the estimates is to evaluate the effect on the target function when one parameter changes keeping all other parameters constant. This is done in *Figure 8* and these plots allow us to confirm that the estimate is at the lowest point and they also give us a visible (but partial) indication of how well identified the different parameters are. In some cases, the parameters end up at corner solutions due to the prior constraints that we have imposed on the steady state (see Section 5.2). Of course, these plots do not tell us what happens to the target function if we change several parameters simultaneously.

The distributions obtained by bootstrapping are shown in *Figure 9*. Convergence evaluation of the distributions shows that the distributions have converged. The parameter estimates and confidence intervals are shown in *Table 2*. Many estimates are uncertain but, as we discuss below, they still give clear indications what kind of model is needed to match the data.

Figure 8. Effects on target function of variations in parameters

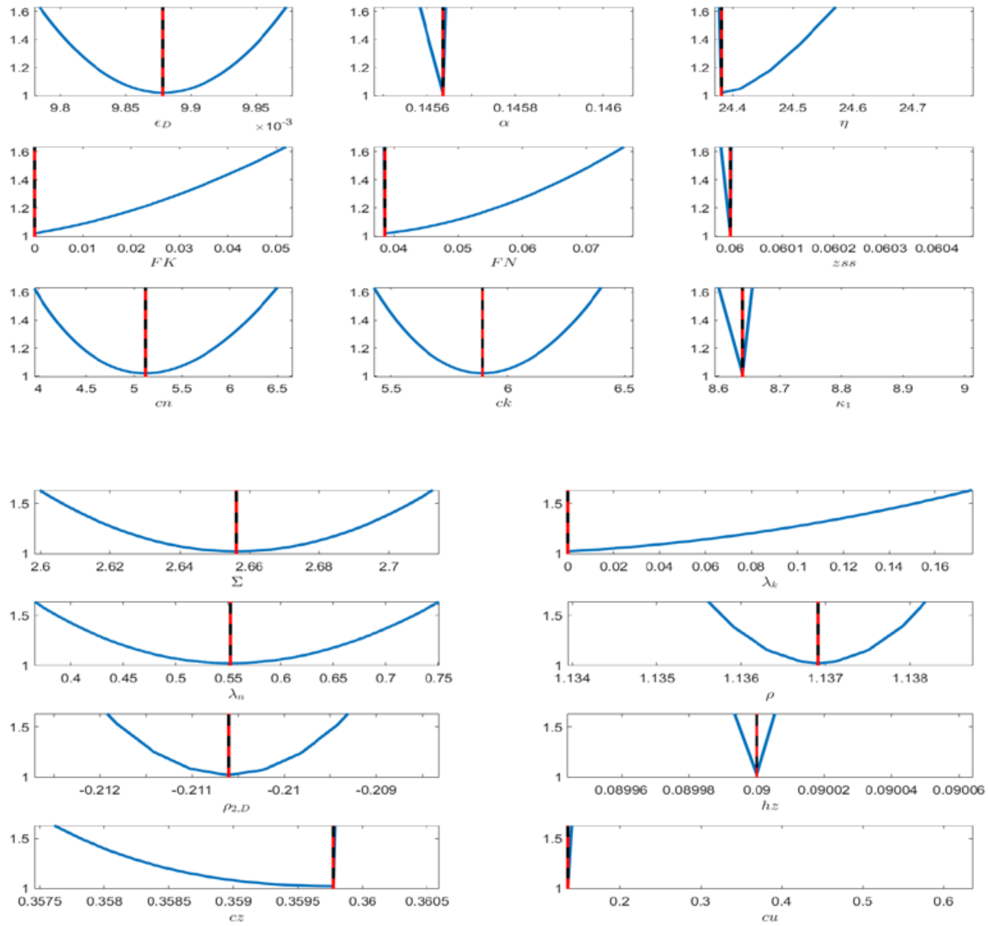
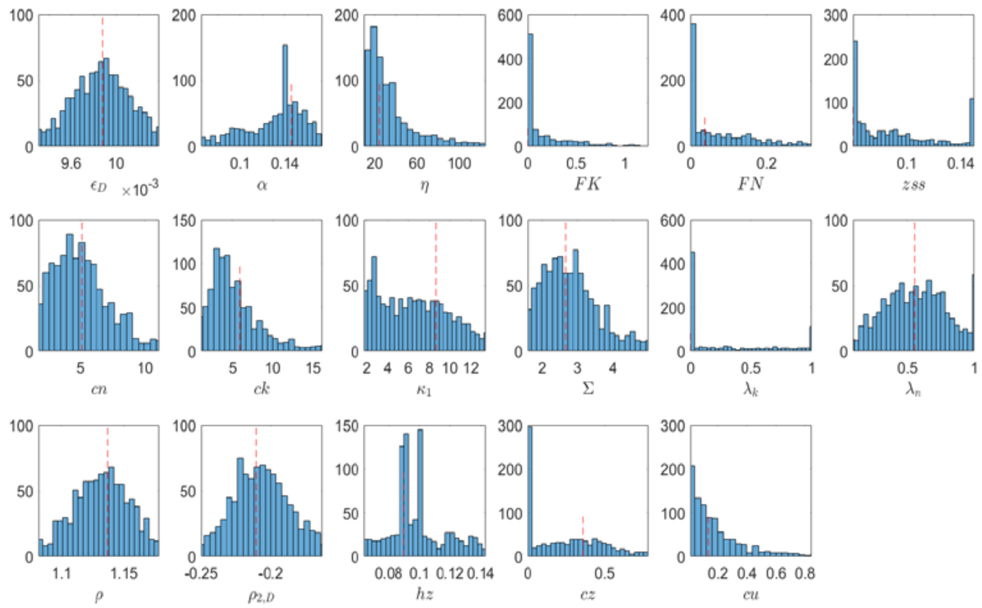


Figure 9. Distributions for deep parameters



Note: 50 of 1000 estimates are outside the intervals shown.

6.4 Discussion of parameter values and their effects on the dynamics

In this section we discuss the parameter estimates and the impact that different parameters have on the dynamics. Throughout, we keep other parameters constant when investigating the effect of changing the value of a particular parameter.[22]

Adjustment costs $\left(c_n = 5.118, c_k = 5.891\right)$

We find substantial and similar adjustment costs for capital and labor and they play a central role. If we set the adjustment costs to zero, this has major effects on the dynamics as almost all the adjustment is made by changing the inputs and there is only a small change in utilization in response to a demand shock.

Implementation lags $\left(\lambda_k = 0, \lambda_n = 0.552\right)$

None of the investment that is made in response to a demand shock is implemented in the same year as the shock occurs. As one would expect, the share of hiring that is carried out in the same year as the demand shock occurs is higher. If we set $\lambda_k = \lambda_n = 1$ so there are no implementation lags, the capital response is speeded up in the first year, but otherwise the responses are very similar to the baseline, so the main role of the implementation lags is to explain the non-response of the capital stock in the first period.

Convexity of the utilization cost $\left(c_u = 0.136\right)$

Our estimates tell us that with substantial adjustment costs and a relatively flat cost function for utilization, firms can meet an increase in demand by telling their workers to work more. If we set $c_u = 2$, so the marginal cost of utilization increases more with u_t, we get a much smaller increase in utilization. Labor and capital still respond sluggishly but the output response is slowed down and becomes much more hump-shaped, peaking 4-5 years after the shock. As a consequence, the inventory stock first declines when demand increases.

[22] With respect to inventories, we chose to hold the steady state level of finished goods' inventories constant so the parameters related to inventories $\left(\kappa_2 \text{ and } \kappa_3\right)$ change when we change some parameter keeping the other structural parameters constant.

Table 2. Parameter estimates and confidence intervals

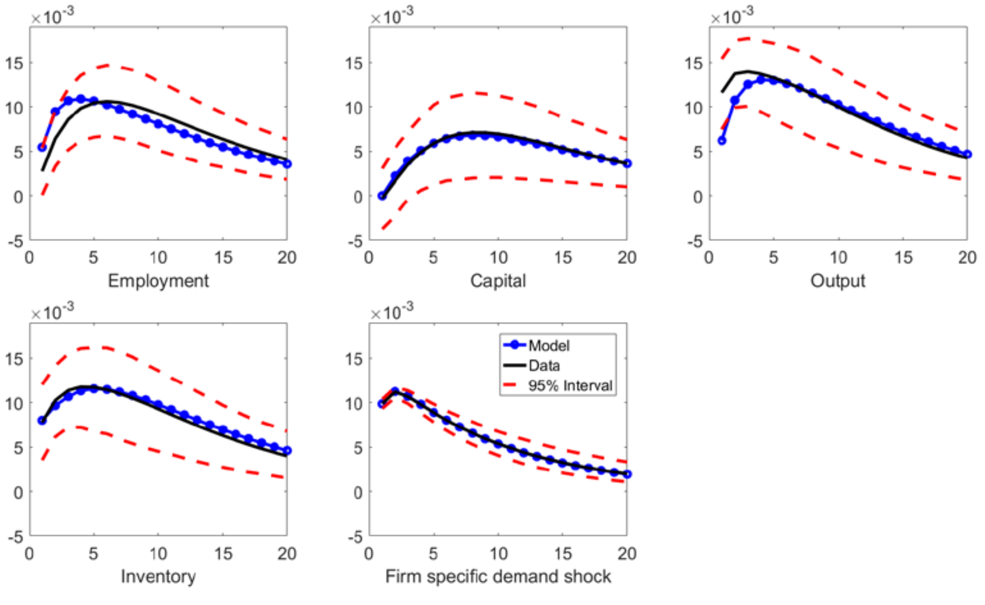
Estimated parameters	**Param.**	**Estimate**	**95 % min**	**95 % max**	**min**	**max**
Std. of the shock	ε	0.010	0.009	0.010	0.00	0.02
Distribution of factor returns	α	0.146	0.054	0.181	0.01	0.99
Demand elasticity	η	24.390	8.662	174.095	2.00	500.00
Fixed capital	FK	0.000	0.000	1.715	0.00	3.00
Fixed employment	FN	0.039	0.000	0.399	0.00	3.00
Finished goods in steady state	zss	0.060	0.060	0.150	0.00	1.00
Labor adjustment cost	cn	5.118	1.351	14.914	0.00	300.00
Capital adjustment	ck	5.891	0.660	20.622	0.00	300.00
Kappa1	kap1	8.643	1.366	14.073	0.20	700.00
Demand Sensitivity	Σ	2.656	1.443	5.809	0.10	6.00
Investment implemented in t	λk	0.000	0.000	1.000	0.00	1.00
Hiring implemented in t	λn	0.552	0.000	1.000	0.00	1.00
AR (t) term demand shock	ρ	1.137	1.069	1.189	-2.00	2.00
AR (t-1) term demand shock	ρ2	-0.211	-0.262	-0.144	-2.00	2.00
Stock of inputs/production	hz	0.090	0.061	0.147	0.00	0.20
Cost of inventories	cz	0.360	0.000	0.893	0.00	2.00
Convexity of utilization cost	cu	0.136	0.010	1.183	0.01	1000.00
Fixed parameters		**Value**				
Elasticity of subs. prod.	σ	0.40				
Inputs/production	m	0.60				
Subjective discount factor	β	0.96				
Price stickiness	θ	0.00				
Depreciation of capital	δk	0.11				
Depreciation of employment	δn	0.11				
Depreciation of inventory	δz	0.11				
Restrictions	**Estimation**	**min**	**max**			
W*N/((P-m)*Y)	0.73	0.73	0.83			
PK*K/((P-m)*Y)	0.79	0.79	0.89			
Z/Y	0.15	0.15	0.25			
hz/zss	1.50	0.67	1.50			
Implied values						
Steady state markup (myss)	0.07					
(S*P-W*N-m*Y-PK*dk*K-cz*ZX) /(P*Y-m*Y)	0.11					

Organizational capital

As it turned out, we did not find an important role for organizational capital. Instead, we could match the impulse-response functions well with flexible utilization. The key difference between costly utilization and a model where the alternative to production work is accumulation of organizational capital lies in the cost of increasing production work today. Utilization is associated with a direct convex cost which may represent compensation for effort and overtime. If, instead, workers spend less time on maintenance and training, this will be costly in terms of future production.

To investigate the difference between the two models, we re-estimated the full model imposing a very steep marginal cost of utilization $\left(c_u = 10000\right)$. The resulting impulse-response functions are shown in *Figure 10*. Now, variations in organizational play an important role but when we eliminate variations in utilization, we fail to match the strong first-period increase in production.

Figure 10. Model with organizational capital instead of utilization

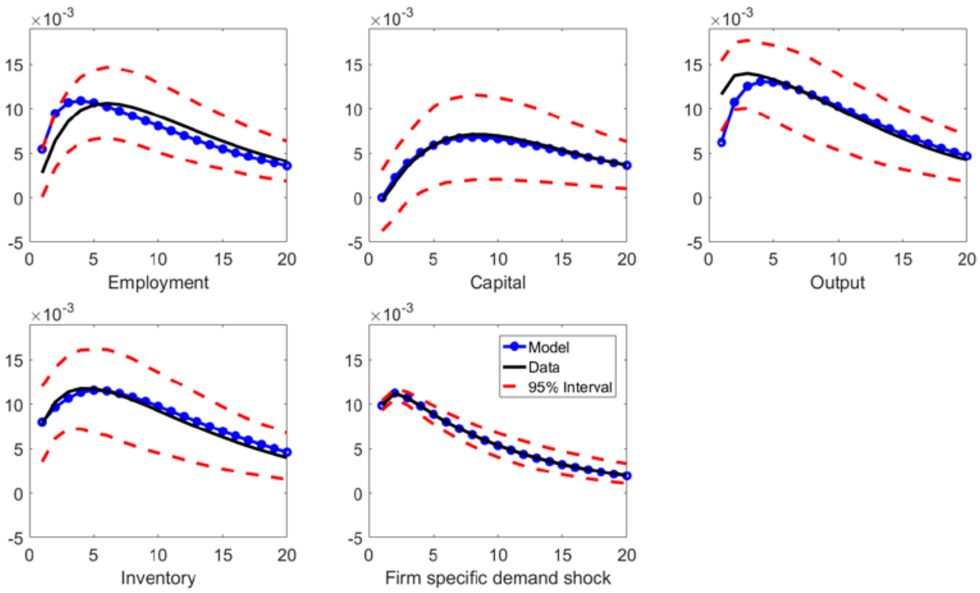


Our interpretation is that this has to do with the hump-shape of the demand shock. The autoregressive parameters in the demand process $\left(\rho_1 = 1.137,\ \rho_2 = -0.211\right)$ imply that the initial increase in demand is followed by a further increase in demand. This makes it unprofitable to de-cumulate organizational capital too much in the first period because this will reduce productivity in the second period, when demand is expected to be even higher. For this reason, a model without variations in effort is unable to match the strong increase in production in the first period and once we allow for variations in effort, we can match the impulse-response functions almost perfectly without variations in organizational capital.

Fixed costs $\left(F_k = 0, F_n = 0.039\right)$

Increasing returns have been proposed as an explanation for why production varies more (in percent) than employment in response to demand-side shocks (Hall (1988)) but we find that increasing returns play little role. If we set $F_k = F_n = 0$, this has negligible effects on the dynamic responses. Furthermore, we find that increasing returns are not sufficient match the impulse-response functions without variations in utilization. This can be seen in *Figure 11*, which shows estimates where we have closed down variations in utilization by setting $\Omega = u = 1$. The empirical IRFs show that, in the first year, the increase in employment is only 23 percent of the increase in output and we would need enormously increasing returns to scale to match this response. At the same time, employment almost catches up with production after a number of years, and this observation is inconsistent with strongly increasing returns to scale.

Figure 11. Model without variations in utilization $\left(u = \Omega = 1\right)$



Distribution of factor returns $(\alpha = 0.146)$

The parameter α affects factor returns but the estimate is difficult to interpret directly when the elasticity of substitution is below unity. Instead, we can look at the cost of labor as a share of value added, which is shown in *Table 2*. It is equal to 73 %, which is the lower bound that we set for the labor share of value added.

Elasticity of substitution $(\sigma = 0.4 \text{ imposed})$

If we increase elasticity of substitution to 1 this has small effects on the dynamics. The elasticity of substitution is poorly identified from these data.

Demand sensitivity $(\Sigma = 2.656)$

According to the estimate, a demand shock of one percent increases demand for the products produced by the typical firm by 2.66 percent. Clearly, our constructed demand variable does capture shocks, which are very important for firms. One reason, why this elasticity is far above unity may be that many firms in manufacturing produce investment goods and durable goods, and that demand for such goods is more sensitive to shocks. Aggregate consumption and investment, which are used to construct the demand variable, contain large portions of services and we know that demand for services is much more stable than demand for manufactured goods. Also, investment in stocks of inputs by other manufacturing firms may respond to the demand shock and contribute to the volatility of demand for the individual firm.

Inventory model $\left(\hat{Z}_{ss} = 0.060, c_z = 0.360, h = 0.090\right)$

In our model there are three costs of holding finished goods inventories: the financing cost, depreciation of the inventory stock $(\delta_z \cdot Z)$ and a storage cost $(c_z \cdot Z)$. We set $\beta = 0.96$ and $\delta_z = 0.11$, we find substantial storage costs $(c_z = 0.360)$ and steady-state inventories end up at the lower bound (15 percent of production). According to the estimates, a relatively large share (2/3) of the inventory stock consists of inputs. This way, the model can match the fact that the inventory response tracks production fairly closely.

Price elasticity and markup $(\eta = 24, \mu_{ss} = 0.07)$

In a model without inventories, a price elasticity of 24 would imply a markup equal to 4 percent but the possibility of stock-out makes the effective demand curve somewhat less elastic for a given inventory stock. A steady state markup of 7 % is in the reasonable range. Christiano et al. (2011) calibrate the markup at 20% "following a wide literature," and Carlsson and Smedsaas (2007) estimate the markup for Swedish manufacturing firms to be 17%.

Discount factor and depreciation rate for capital $(\beta = 0.96,\ \delta_k = 0.11 \text{ imposed})$

We had difficulty estimating these parameters so we set them at reasonable values. If we change the discount factor to 0.92 or the depreciation rate for capital to 7 percent this has little effect on the dynamic responses. These parameters are poorly identified from these data.

Separation rate for labor $(\delta_n = 0.11 \text{ imposed})$

The parameter δ_n can be interpreted as an exogenous separation rate and then we would expect a value around 0.10. There is, however, an alternative interpretation, and that is that the convex adjustment cost arises when a firm *changes the number of workers* instead of when it hires and this would imply $\delta_n = 0$. This is a reasonable interpretation since a change in the number of workers means that tasks must be reallocated between the workers and new workplaces must be arranged. If we set $\delta_n = 0$ this has very small effects on the responses, so it appears that this parameter is not well identified with the data that we have.

Depreciation rate for inventories $(\delta_z = 0.11 \text{ imposed})$

The depreciation rate for inventories was also set to 11 percent. If we increase it to 20 percent, this has small effects on the dynamics. If we decrease it to 0.05, this affects the dynamics but we view such a depreciation rate as implausibly low. Technical change and changes in fashion and design may make goods unsellable, so the depreciation rate of finished goods should be relatively high.

Price stickiness and variations in the markup $(\theta = 0 \text{ imposed})$

We set $\theta = 0$, making prices completely flexible. Still there is a very weak price response to the demand shock and this explains why we could not estimate the degree of price rigidity with any precision. The weak price response is due to utilization being very flexible, so marginal cost increases only slightly in response to the demand shock. The finding of a flat marginal cost curve is consistent with Carlsson and Nordström-Skans (2012).

Another factor, which contributes to a weak price response, is that the markup declines somewhat in response to an increase in demand (*Figure 7*). This can be understood by looking at equation (23) which determines sales. As noted above, the possibility of stock-outs makes demand less price sensitive than it would be without stock-outs. Furthermore, the price sensitivity of sales depends on the level of demand compared to the inventory stock. In the first period, demand increases while inventories of finished goods remain roughly unchanged and, as a result, the probability of stock-out decreases. This makes demand more sensitive to the price, and hence the firm reduces its markup. Note, however, that we are not using price data in the estimation, so our inference about prices is very indirect.

7. Relation to previous research

Increasing returns vs. factor utilization

Hall (1988) showed that shocks, which should be uncorrelated with technology are associated with variations in output that are more than proportional to the corresponding variation in inputs and he interpreted this as evidence of increasing returns to scale. With increasing returns, firms make losses if price is equal to marginal cost and since firms typically do not make losses, not even in periods of low demand, Hall concluded that firms must have very substantial market power.

Increasing returns may take the form of increasing returns in the long run production function or there may be some form of *short run increasing returns* because some factors of production are *quasi-fixed*. Hall (1988) discusses a case where some predetermined amount of overhead labor determines the firm's maximum production capacity and a fixed amount of production labor is needed per unit of output actually produced. This means that overhead labor is a fixed cost in the short run and when the firm operates below capacity, the marginal cost is the cost of the required production work; again the firm would make a loss if price would be equal to marginal cost. Importantly, Hall assumes that there is no cost of increasing the utilization of overhead labor.

An alternative explanation is that there are variations in utilization (e.g. effort) so the services of labor and capital vary more than measured inputs, and that variations in utilization are *costly*. Burnside, Eichenbaum and Rebelo (1993) found that a model with constant returns to scale, perfect competition, implementation lags in employment and variations in effort fits the data well and that it can account for a positive correlation between the growth rates of the Solow residual and government expenditure.[23] Note that, if the cost of increasing utilization is not properly accounted for, the marginal cost will be underestimated.

Our identification strategy is conceptually similar to the one used by Hall (1988). In line with Hall, we find that firms have market power but increasing returns are not found to be important for medium-term dynamics. Instead, costly variations in utilization play an important role. Note, however, that we do not have data for hours worked – only the reported number of “full-time equivalent employees” which probably do not fully reflect variations in hours. This means that we do not know how much of the variation in utilization that is variations in unregistered work hours or effort and how much is variations in registered overtime. In this respect, our results are not directly comparable to Hall’s results.

Like Burnside, Eichenbaum and Rebelo (1993), we find that adjustment lags and variations in utilization are important, but our dynamic specification is quite different. Their model has no adjustment costs but employment is determined one period in advance, leading to variations in utilization with very low persistence. According to our estimates, there is only a small implementation lag in employment, but adjustment costs lead to very sluggish adjustment and large and persistent variations in utilization. In this respect, our results are similar to those of Fairise and Langot (1994) and Braun and Evans (1998).

Imbs (1999) adjusted Solow residuals for variations in utilization of capital and labor and he found that the adjusted residuals were substantially less pro-cyclical than standard series. In his model, utilization can be backed out due to specific functional forms for the utilization costs.[24] Our utilization cost function is more general, so our results are more data-driven and less dependent on theoretical assumptions, but the conclusions are similar.

[23] In both papers, the term *labor hoarding* is used for a model where labor is somehow quasi-fixed, but while Hall assumes that utilization of the quasi-fixed labor is costless, Burnside, Eichenbaum and Rebelo include a convex cost of utilization.

[24] In the model of Imbs (1999), the cost functions for utilization of capital and labor have only one free parameter so that parameter can be backed out from the steady-state conditions. This is not the case in our model.

Straight time and overtime

Lucas (1970), Sargent (1987) and Hansen and Sargent (1988) argued that imperfect substitution between straight time and overtime can help to explain the procyclical pattern of the standard Solow residual. Sargent (1987) writes the production function $Y_t = A_t K_t^{1-\alpha} n_{1t}^{\alpha} + A_t K_t^{1-\alpha} n_{2t}^{\alpha}$ where n_{1t} is straight-time and n_{2t} is overtime. This production function is a special case of our production function. To see this, rewrite it as $Y_t = A_t K_t^{1-\alpha} n_{1t}^{\alpha} \left(1+\left(n_{2t} / n_{1t}\right)^{\alpha}\right)$. Setting $n_{1t} = N_t$ and $1+\left(n_{2t} / n_{1t}\right)^{\alpha} = u_t$ we get our production function with the elasticity of substitution is set to one.[25]

Time spent on maintenance, cleaning and training

There is ample evidence that workers spend a substantial fraction of their time on tasks that increase future rather than current production; see e.g. Fay and Medoff (1985) and Kim and Lee (2007). Kim and Lee showed in a theoretical model that, even without adjustment costs, skill accumulation will be countercyclical in a real business cycle model because the opportunity cost of skill accumulation is higher when productivity is high. Similar ideas have emerged in the growth literature: Aghion and Saint-Paul (1998) consider a model where productivity growth is costly in terms of current production and where productivity growth turns out to be countercyclical.[26]

We incorporated these ideas in our model by assuming that workers spend some of their time building “organizational capital” that increases future productivity. However, our estimates did not support the idea that firms invest less in organizational capital when there is high demand. The model with a relatively flat cost of utilization does a better job matching the impulse-response functions.

Price rigidity

Rotemberg and Summers (1990) argued that price rigidity can explain pro-cyclical productivity under perfect competition. They assume that firms must fix prices before demand is known. With free entry, price must be equal to average cost, so the price will be above marginal cost in a recession. Rotemberg and Summers assume that firms produce at the point

[25] Hall (1996) estimated a model with straight time and overtime on macro data and found that such a model delivers greater magnification and propagation of shocks than the model by Burnside, Eichenbaum and Rebelo. In our context, these models appear observationally equivalent.

[26] See also DeJong and Ingram (2001). Cooper and Johri (2002) assume, instead, that there is learning-by-doing, so the accumulation of “organizational capital” is positively related to the level of production.

where marginal cost equals the predetermined price in a boom, so there is rationing in periods of high demand. This implies that, although there is perfect competition, the price will *on average* be higher than marginal cost and this could explain the results found by Hall (1988).

In our model, the markup is sufficiently large so that price exceeds marginal cost throughout the adjustment to the shock and firms always want to satisfy demand. However, a positive demand shock makes demand high relative to the stock of finished goods, so there is increased rationing in the sense that a larger fraction of the customers do not find their desired variety.

Adjustment costs

We find substantial convex adjustment costs and they play a key role explaining the dynamics. We are fully aware that quadratic adjustment costs are a crude approximation. Many authors have found evidence of asymmetric, linear or lumpy adjustment costs (for reviews, see Hamermesh and Pfann (1996), Adda and Cooper 2003). Note, however, that our model is meant to capture the *average* reaction of firms to demand shocks and we may succeed in doing that even if adjustments by individual firms are lumpy. In fact, we manage to fit average responses very well with convex adjustment costs.[27]

Marshallian externalities

Marshallian externalities have been considered by many authors; see Cooper and Haltiwanger (1996) and Braun and Evans (1998) for references. Typically, such externalities are modeled by assuming that aggregate output affects the productivity of the individual firm. Since we consider effects of firm-specific shocks, aggregate externalities are not relevant for our results. Any such effects will be picked up by the time dummies.[28]

Inventory dynamics

As emphasized by Bils and Kahn (2000), inventory behavior provides clues to the nature of business cycles. Some researchers have viewed pro-cyclical inventory investment as evidence that the costs of producing must be low in boom periods. For example, Khan and Thomas (2007) interpret pro-cyclical inventory investments as resulting from supply-side shocks. Our

[27] Our adjustment costs could also represent search frictions.

[28] Our constructed demand shocks are firm specific because firms have different export shares. Still, much of the variation that remains after time effects is industry-specific and one can imagine Marshallian externalities on the industry level. Unfortunately, we do not see how such externalities could be separated empirically from industry-wide shocks.

study says nothing about the role of productivity shocks, or how they affect inventory holdings, but we find a very strong positive response of inventory holdings to demand shocks, which is well explained by a model with a stock-out motive, as suggested by Kahn (1987, 1992). Other studies finding support for a model stock-out motive are Bils and Kahn (2000), Wen (2005) and Kryvtsov and Midrigan (2013). [29]

While inventory investments respond positively to demand shocks, they fail to keep pace with shipments: the *ratio* of finished-goods inventories to production (and sales) decreases when demand increases. To understand how this is explained in the model, note that the first order condition for finished-goods inventory holdings is

$$v_t + m - \beta\left(1-\delta_z\right)E_t\left(v_{t+1}+m\right)+c_z = \left(\kappa_1 - 2\kappa_3 \frac{\hat{Z}_t}{D_t^{\Sigma}P_t^{-\eta}}\right)\left(P_t - v_t - m\right). \tag{41}$$

The left hand side is the cost of holding an additional unit of finished goods inventories: the marginal cost today minus the expected discounted marginal cost next year plus the storage cost. The right hand side is the effect of inventories on sales (due to reduced stock-outs) times the value of selling one more unit (the markup). In a model without stock-outs, inventories have no effect on sales and the right hand side is zero. This leads to pure production-smoothing: inventories are adjusted until marginal cost today is equal to discounted marginal cost next year, taking account of the storage cost. When inventories contribute to sales, the desired ratio of inventories to demand depends on prices. As pointed out by Bils and Kahn (2000), at least one of the following things must happen for the ratio of inventories to sales (and demand) to decrease when there is an increase in demand:

1. marginal costs increase relative to discounted future marginal costs, or
2. the markup is countercyclical.

Bils and Kahn argued that there is little evidence of predictable changes in marginal costs but that the markup is lower in booms. In line with Bils and Kahn (2000) we find that the markup decreases when there is an increase in demand and there is also a predictable decline in the marginal cost, so both mechanisms are at work in our model. However, we should note three caveats here. First, we do not use price data because we do not have micro price data for all the firms. Second, a more realistic modeling of the demand side and financial conditions may lead to a more countercyclical markup; see Gottfries (1991), Rotemberg and Woodford (1999)

[29] In a closely related approach, Kydland and Prescott (1982), Christiano (1988) and Ramey (1989) introduce inventories as a factor of production.

and Lundin et al. (2009). Third, we identify responses to shocks that are more persistent than normal business cycle fluctuations.

Permanent and transitory shocks

Our autoregressive model allows for a hump-shaped and persistent response to the demand shock but we do not allow for unit roots. An alternative would be to allow for permanent and transitory shocks as is done by Franco and Philippon (2007) and Carlsson, Messina and Nordström Skans (2017) but we leave this for future research. Still, the high persistence of the demand shock is in line with results of Franco and Philippon (2007) who found that four fifths of the variation on the firm level is driven by persistent shocks to technology and the composition of demand while much of the common variation across firms is due to temporary macroeconomic shocks. Carlsson, Messina and Nordström Skans (2017) use long run restrictions to identify demand shocks and they also find that demand shocks have big effects on production, employment and labor productivity. They find that the employment response to permanent demand shocks is quick while the response to temporary shocks is more muted. In our specification, the demand shock is persistent but not permanent and impulse-response function shows a sluggish response to the demand shock.

8. Conclusion

Investment, hours worked, labor productivity and inventory holdings are all pro-cyclical but there is no consensus on how to interpret these correlations. A positive correlation between production and output per worker may arise because of productivity shocks drive both variables or because demand-side shocks lead to variations in factor utilization. A positive correlation between production and inventory holdings may arise because firms invest in inventories when productivity is high or because you need more inventory holdings in order not to stock out when demand is high. Since all variables are endogenous on the macroeconomic level, it is hard to establish causality without additional assumptions about functional forms and the stochastic nature of the shocks as is done in DSGE models.

The same problem occurs if we use panel data for individual firms: a positive correlation between production and output per worker may be interpreted in different ways and without some exogenous source of variation it is hard to establish causality.

In this paper, we tried to study causal effects of demand-side shocks on firms' decisions. We combined macro and micro data using input-output tables so as to identify demand-side shocks which are exogenous for individual firms and we used firm-level panel data to study how individual firms react to such shocks. Then we used this information to estimate the deep structural parameters of a theoretical model. We found that registered inputs of labor and capital respond slowly to an increase in demand while production and inventory holdings respond quickly. These responses can be well explained by a theoretical model with adjustment costs, implementation lags (time to build), variable utilization of the production factors and a stock-outs-avoidance motive for inventories.

Since we only study the effects of specific demand shocks, we cannot say anything about the effects of other shocks or the relative importance of supply and demand shocks for business cycle fluctuations. Still, our estimates may be useful as reference points for researchers who are estimating or calibrating DSGE models of the business cycle.

It is worth emphasizing that our methodology is fundamentally different from the one used by many studies, which estimate general equilibrium models. In a general equilibrium model, issues having to do with labor supply (intertemporal substitution and income effects on labor supply) play central roles. In our estimation, any such general equilibrium effects are picked up by the time dummies. We assume that each individual firm can hire as much labor and buy as much capital as it desires at given prices, which may vary over time. We ask how an individual firm reacts to a *firm-specific* shock. This makes our analysis more limited and partial, but also more focused. By improving our understanding firms' behavior we hope to shed light on some of the building blocks of macroeconomic models.

References

Ackerberg, Daniel A., Kevin Caves and Garth Frazer, 2015, Identification Properties of Recent Production Function Estimators, *Econometrica*, 83, 2411-2451.

Adda, Jérome and Russel Cooper, 2003, *Dynamic Economics – Quantitative Methods and Applications*, MIT Press , Cambridge, Massachusetts.

Aghion, Philippe, and Gilles Saint-Paul, 1998, Virtues of Bad Times, *Macroeconomic Dynamics*, 2, 322-344.

Arellano, Manuel and Stephen Bond, 1991, Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *The Review of Economic Studies*, 58, 277–297.

Baily, Martin N., Eric J.Bartelsman and John Haltiwanger, 2001, Labor Productivity: Structural Change and Cyclical Dynamics, *Review of Economics and Statistics* 83, 420-433.

Basu, Susanto, 1996, Procyclical Productivity: Increasing Returns or Cyclical Utilization?, *Quarterly Journal of Economics,* 111, 719-751.

Basu, Susanto, and John G. Fernald, 1997, Returns to Scale in U.S. Production: Estimates and Implications, *Journal of Political Economy*, 105, 249-283.

Bils, Mark and James A. Kahn, 2000, What Inventory Behavior Tells Us about Business Cycles, *American Economic Review*, 90, 458-481.

Braun, R. Anton, and Charles Evans, 1988, Seasonal Solow Residuals and Christmas: A Case for Labor Hoarding and Increasing Returns, *Journal of Money, Credit and Banking*, 30, 306-330.

Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo, 1993, Labor Hoarding and the Business Cycle, *Journal of Political Economy*, 101, 245–273.

Carlsson, Mikael, Stefan Eriksson, and Nils Gottfries, 2013, Product Market Imperfections and Employment Dynamics, *Oxford Economic Papers*, 65, 447–470.

Carlsson, Mikael, Julián Messina and Oskar Nordström-Skans, 2017, Firm-Level Shocks and Labor Flows, mimeo, Uppsala University.

Carlsson, Mikael and Oskar Nordström-Skans, 2012, Evaluating Microfoundations for Aggregate Price Rigidities: Evidence from Matched Firm-Level Data on Product Prices and Unit Labor Cost, *American Economic Review*, 102, 1571-1595.

Carlsson, Mikael and Jon Smedsaas, 2007, Technology Shocks and the Labor-Input Response: Evidence from Firm-Level Data, *Journal of Money, Credit and Banking*, 39, 1509–1520.

Chirinko, Robert S., Steven M. Fazzari, and Andrew P. Meyer, 2011, A New Approach to Estimating Production Function Parameters: The Elusive Capital–Labor Substitution Elasticity, *Journal of Business & Economic Statistics*, 29, 587–594.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans, 2005, Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy, *Journal of Political Economy*, 113, 1–45.

Cooper, Russel, and Alok Johri, 2002, Learning-by-doing and Aggregate Fluctuations, *Journal of Monetary Economics*, 49, 1539-1566.

DeJong, David and Beth F. Ingram, 2001, The Cyclical Behavior of Skill Acquisition, *Review of Economic Dynamics*, 4, 536-561.

Eriksson, Stefan and Karolina Stadin, 2017, What are the Determinants of Hiring? The Importance of Product Market Demand and Search Frictions, Applied Economics, 49, 5144-5165.

Fairise, Xavier and Francois Langot, 1994, Labor Productivity and the Business Cycle: Can R.B.C Models be Saved?, *European Economic Review*, 38, 1581-1594.

Fay, Jon A. and James L. Medoff, 1985, Labor and Output over the Business Cycle: Some Direct Evidence, *American Economic Review*, 75, 638-655.

Field, Alexander J., 2010, The Procyclical Behavior of Total Factor Productivity in the United States, 1890-2004, *Journal of Economic History*, 70, 326-350.

Franco, Francesco and Thomas Philippon, 2007, Firms and Aggregate Dynamics, *Review of Economics and Statistics*, 89, 587–600.

Gertler, Mark and Peter Karadi, 2015, Monetary Policy Surprises, Credit Costs, and Economic Activity, *American Economic Journal: Macroeconomics*, 7, 44-76.

Greenwood, Jeremy, Zvi Hercowitz and Per Krusell, 2000, The Role of Investment-Specific Technological Change in the Business Cycle, 44, 91-115.

Hall, Robert E., 1988, The Relation between Price and Marginal Cost in U.S. Industry, *Journal of Political Economy*, 96, 921-947.

Hamermesh, Daniel S. and Gerard Pfann, 1996, Adjustment Costs in Factor Demand, *Journal of Economic Literature*, 34, 1264-1292.

Hart, Robert A. and James R. Malley, 1999, Procyclical Labor Productivity: A Closer Look at a Stylized Fact, *Economica*, 66, 533-550.

Imbs, Jean M., 1999, Technology, Growth, and the Business Cycle, *Journal of Monetary Economics*, 44, 65-80.

Kahn, James A., 1987, Inventories and the Volatility of Production, *American Economic Review*, 77, 667-679.

Kahn, James A., 1992, Why is Production More Volatile than Sales? Theory and Evidence on the Stockout-Avoidance Motive for Inventory-Holding, *Quarterly Journal of Economics*, 107, 481-510.

Khan, Aubhik and Julia K. Thomas, 2007, Inventories and the Business Cycle: An Equilibrium Analysis of (S,s) Policies, *American Economic Review*, 97, 1165-1188.

Kim, Daehaeng, and Chul-In Lee, 2007, On-the-job Human Capital Accumulation in a Real Business Cycle Model: Implications for Intertemporal Substitution Elasticity and Labor Hoarding, *Review of Economic Dynamics*, 10, 494-518.

Kryvtsov, Oleksiy, and Virgiliu Midrigan, 2013, Inventories, Markups, and Real Rigidities in Menu Cost Models, *Review of Economic Studies*, 80, 249-276.

Kydland, Finn and Edward Prescott, 1982, Time to Build and Aggregate Fluctuations, *Econometrica*, 50, 1345-1370.

Levinsohn, James and Amil Petrin, 2003, Estimating Production Functions Using Inputs to Control for Unobservables, *Review of Economic Studies*, 70, 317-341.

Mickelsson, Glenn, 2016, Estimation of DSGE Models with Uninformative Priors, in *DSGE Model Estimation and Labor Market Dynamics*, PhD Thesis, Economic Studies 163, Department of Economics, Uppsala University.

Nadiri, M. Ishaq and Ingmar R. Prucha, 1996, Estimation of the Depreciation Rate of Physical and R&d Capital in the U.S. Total Manufacturing Sector, *Economic Inquiry*, 34, 43–56.

Nelder, J. A. and R. Mead, 1965, A Simplex Method for Function Minimization, *The Computer Journal*, 7, 308–313.

Olley, Steven and Ariel Pakes, 1996, The Dynamics of Productivity in the Telecommunications Equipment Industry, *Econometrica*, 64, 1263-1295.

Ramey, Valerie A., 1989, Inventories as Factors of Production and Economic Fluctuations, *The American Economic Review*, 79, 338-354.

Ramey, Valerie A., 2016, Macroeconomic Shocks and Their Propagation, *Handbook of Macroeconomics*, 2, 71-162.

Smets, Frank and Rafael Wouters, 2007, Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach, *The American Economic Review*, 97, 586–606.

Stadin, Karolina, 2014, *Employment Dynamics*, PhD Thesis, Economic Studies 145, Department of Economics, Uppsala University.

Stadin, Karolina, 2015, Firms' Employment Dynamics and the State of the Labor Market, Working Paper Series 2015:20, IFAU.

Wen, Yi, 2005, Understanding the Inventory Cycle, *Journal of Monetary Economics*, 52, 1533-1555.

APPENDIX

Construction of the steady state

The estimation involves a very large number of repeated simulations of the model. To save time in this process, we calculate the steady state values analytically instead of searching for the steady state. Let variables with a bar denote steady-state values. Without loss of generality we can chose units so that

$$\bar{Y} = \bar{K} = \bar{N} = \bar{\Omega} = \bar{P} = \bar{D} = \bar{u} = 1\,,\ \bar{I} = \delta_k\,,\ \bar{H} = \delta_n\,,\ \bar{q} = \bar{\gamma} = 0. \tag{42}$$

For given values of δ_ω and χ we can calculate x in steady state: $\bar{x} = 1 - \delta_\omega / \chi$ and then our normalizations imply a value for *A*:

$$A = \left(\alpha + (1-\alpha)\,\bar{x}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{1-\sigma}}. \tag{43}$$

It is convenient to view the steady state inventory stock of finished goods as a parameter to be estimated. Denoting this value $\bar{Z}$ we get steady state sales as $\bar{S} = 1 - \delta_z \bar{Z}$. In the steady state we have

$$\kappa_1 \bar{Z} + \kappa_2 - \kappa_3 \bar{Z}^2 = \bar{S} \qquad \frac{\bar{S}/\eta}{1-\bar{v}-m} = \kappa_2 + \kappa_3 \bar{Z}^2$$

$$\frac{r_z\left(\bar{v}+m\right)+c_z g_t}{1-\bar{v}-m} = \kappa_1 - 2\kappa_3 \bar{Z} \qquad \text{where } r_z = 1+\beta\left(1-\delta_z\right).$$

Multiplying the last equation by $\bar{Z}$ and summing both sides of these equations we can solve for the marginal cost of real value added in steady state:

$$\bar{v} = \frac{1-m-1/\eta-\left(r_z m + c_z\right)\bar{Z}/\bar{S}}{1+r_z\bar{Z}/\bar{S}}\ .$$

For a given estimate of κ_1 we can then use the equations above to solve for κ_3 and κ_2.

We can also find the capital price and wage that are consistent with our normalizations:

$$P^K = \beta\bar{v}\alpha / \left(1-\beta\left(1-\delta_K\right)\right) \qquad \text{and} \qquad W = \bar{v}\left(1-\alpha\right)\bar{x}^{\frac{\sigma-1}{\sigma}}.$$

The first-order conditions for x and Ω give us

$$a = \frac{\left(1-\beta\left(1-\delta_\omega\right)\right)\left(1-\alpha\right)}{\beta\xi\chi\bar{x}^{1/\sigma}} + 1.$$

Furthermore, the normalizations imply $\Phi_u = \bar{v}$ and $\bar{\phi} = \beta\bar{v}\left(a-1\right)\xi / \left(1-\beta\left(1-\delta_\omega\right)\right)$.

Table A1. Industries used in the estimations and the number of firms in each industry in the balanced panel, all in the manufacturing sector

Sector (SNI 92)	Number of firms
15 Food products and beverages	31
16 Tobacco products	0
17 Textiles	13
18 Wearing apparel; furs	2
19 Leather and leather products	4
20 Wood and products of wood and cork (except furniture); articles of straw and plaiting materials	53
21 Pulp, paper, and paper products	32
22 Printed matter and recorded media	17
23 Coke, refined petroleum products and nuclear fuels	1
24 Chemicals, chemical products, and man-made fibers	30
25 Rubber and plastic products	65
26 Other non-metallic mineral products	15
27 Basic metals	18
28 Fabricated metal products, except machinery and equipment	189
29 Machinery and equipment n.e.c.	168
30 Office machinery and computers	2
31 Electrical machinery and apparatus n.e.c.	43
32 Radio, television, and communication equipment and apparatus	20
33 Medical, precision and optical instruments, watches, and clocks	24
34 Motor vehicles, trailers, and semi-trailers	53
35 Other transport equipment	8
36 Furniture; other manufactured goods n.e.c.	50
Total number of firms	838

Note: Industries are defined according to SNI 92 and SNI2002 are almost the same on the 2-digit level.

Table A2. Estimated empirical model

Equations for firm level production, employment, capital stock, and inventories

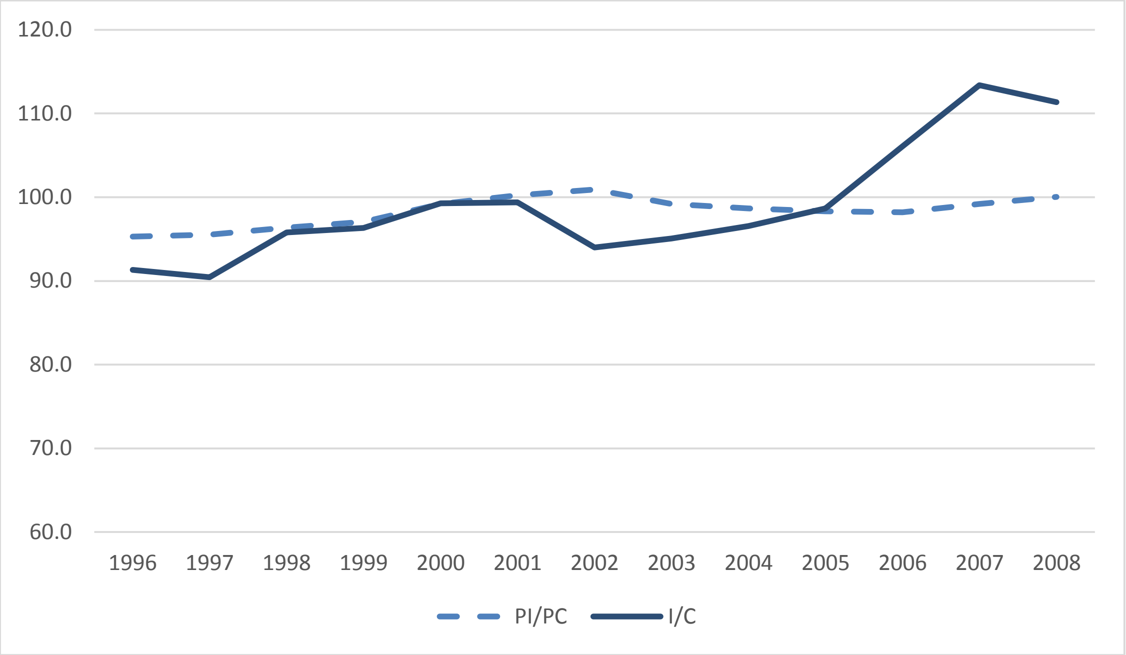
	(1) lnYr	(2) lnN	(3) lnKr	(4) lnZr
L.lnYr	0.550***	0.134***	0.101***	0.185***
	(0.020)	(0.013)	(0.014)	(0.021)
L2.lnYr	0.024	-0.034***	-0.021	0.015
	(0.018)	(0.011)	(0.015)	(0.020)
L.lnN	0.117***	0.658***	0.059***	0.112***
	(0.021)	(0.029)	(0.020)	(0.028)
L2.lnN	-0.015	0.018	0.005	-0.019
	(0.020)	(0.020)	(0.021)	(0.024)
L.lnKr	0.039***	0.056***	0.821***	0.018
	(0.014)	(0.011)	(0.019)	(0.019)
L2.lnKr	-0.033**	-0.047***	-0.076***	-0.018
	(0.014)	(0.009)	(0.016)	(0.017)
L.lnZr	0.148***	0.072***	-0.002	0.491***
	(0.013)	(0.009)	(0.010)	(0.022)
L2.lnZr	-0.057***	-0.033***	-0.017*	0.005
	(0.012)	(0.008)	(0.010)	(0.016)
lnD	1.173***	0.277**	-0.031	0.787***
	(0.207)	(0.139)	(0.175)	(0.218)
L.lnD	-0.737***	-0.059	0.094	-0.490**
	(0.205)	(0.135)	(0.184)	(0.229)
Observations	8,143	8,143	8,143	8,143
Number of FAD_F_Id	818	818	818	818
R-squared	0.572	0.651	0.673	0.505

AR(2) process for product demand

	lnD
L.lnD	1.135***
	(0.022)
L2.lnD	-0.208***
	(0.020)
Observations	9,169
Number of FAD_F_Id	838
R-squared	0.994
St.D. of residual	0.0099

Note: Robust standard errors in parentheses, clustered at the firm level; *** p<0.01, ** p<0.05, * p<0.1. Time dummies and firm fixed effects are included in all regressions. Balanced panel of firms in manufacturing with at least 10 employees in Sweden 1996-2008. Firms with extreme values removed. The estimation method is OLS. ; We also tried to do diff-GMM estimation, but we were not able to find an instrument set which is both valid and enough relevant to give a good identification.

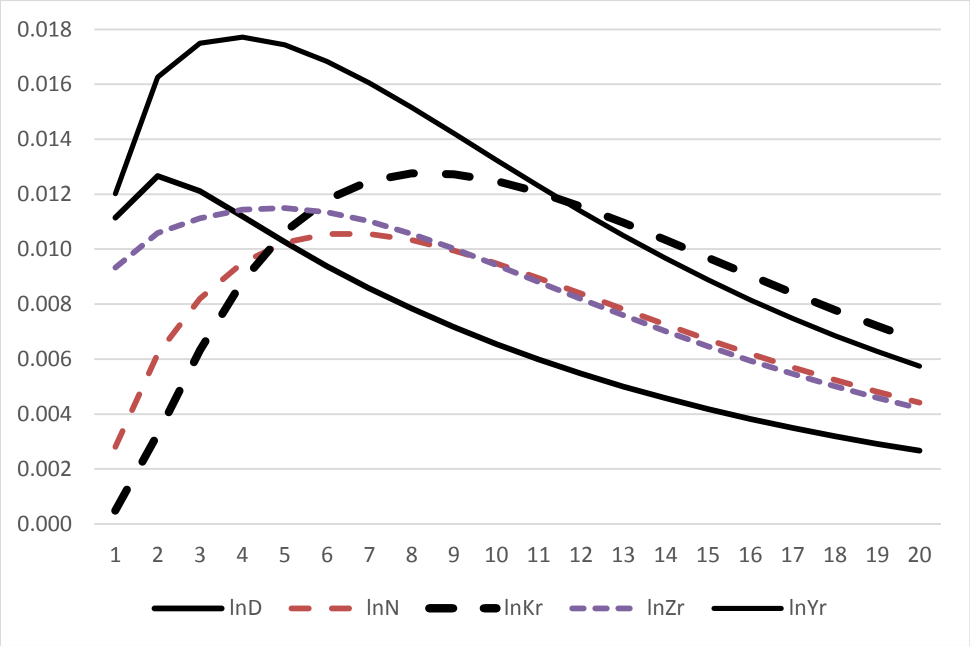
Figure A1. The relative price of investment goods



Note: I/C is the ratio of investment to consumption, (volume indexes) and PI/PC is the ratio of the corresponding deflators

Figure A2. Larger and smaller firms

Larger firms (mean N ≥ 50)



Smaller firms (mean N < 50)