Unraveling Firms: Demand, Productivity and Markups Heterogeneity*

Emanuele Forlani† Ralf Martin‡ Giordano Mion§ Mirabelle Muûls¶

First draft: 16th June 2015. This draft: 17th June 2015.

Abstract

We develop a new econometric framework simultaneously allowing for heterogeneity in demand, TFP and markups across firms while leaving the correlation among the three unrestricted. We are able to do this by systematically exploiting many of the assumptions that are implicit in most previous firm-level productivity estimation approaches. We use production data on Belgian firms to quantify TFP, demand and markups and show how our measures are correlated among them, across time as well as with measures obtained from other approaches. We also show how and to what extent our three dimensions of heterogeneity allow to gain deeper and sharper insights on the understanding of two key firm outcomes: export status and size.

PRELIMINARY AND INCOMPLETE: PLEASE DO NOT QUOTE OR CITE WITHOUT PERMISSION

Keywords: Demand heterogeneity, Productivity heterogeneity, Markups Heterogeneity, Monopolistic competition, Export status, Firm size

JEL Classification: D24, L11, L25, F14

---

* We thank Andrea Ariu, Andrew Bernard, Jan De Loecker and Ilke Van Beveren. Most of this research has been carried out while Giordano Mion was affiliated to the National Bank of Belgium. The views expressed in this paper are our own and do not necessarily reflect those of the National Bank of Belgium.

† Department of Economics and Management, University of Pavia, Pavia, Italy; emanuele.forlani@eco.unipv.it; website: https://sites.google.com/site/emanueleforlani79/home. Also affiliated with the Centro Studi Luca d’Agliano.

‡ Imperial College Business School, Imperial College, London, UK; r.martin@imperial.ac.uk; website: http://www.imperial.ac.uk/people/r.martin. Also affiliated with the Centre for Economic Performance (CEP - London School of Economics) and the Grantham Institute.

§ Surrey Business School, University of Surrey, Guildford, UK; g.mion@surrey.ac.uk; website: http://surrey.ac.uk/sbs/people/giordanomion. Also affiliated with the CEP and the Centre for Economic Policy Research.

¶ Imperial College Business School and Grantham Institute, Imperial College, London, UK; m.muuls@imperial.ac.uk; website: http://www.imperial.ac.uk/people/m.muuls. Also affiliated with the CEP.
1 Introduction

Economists have an interest in estimating firm-level productivity in a range of sub-disciplines. These estimates are indeed often used as inputs in a number of wide ranging applications concerning the firm size distribution, firm survival and growth, self-selection of firms into export status and the extensive and intensive margins of trade to name a few. This paper proposes a new framework for firm level productivity estimation simultaneously allowing for heterogeneity in technical efficiency, demand, and markups across firms while leaving the correlation among the three unrestricted.

In the literature, the most commonly used approach involves estimating a production function by regressing output quantity on input quantity and using the resulting residual shock as a productivity index typically referred to as Total Factor Productivity (TFP). This raises three issues. First, most studies do not have output quantity data available at the firm-level so that regressions are fitted using revenue data, i.e., price times quantity. Second, a well known issue is the endogeneity of production factors used as explanatory variables. Third, in addition to technical efficiency, firms could be heterogeneous in other dimensions. For example, vertical differentiation may lead firms to face rather different demand functions. At the same time market power variations, due to product quality or technical efficiency differences between firms, could also substantially affect the size of the markup over the marginal costs that firms manage to extract from customers. Last but not least, it is also possible that there are further differences in market power that are not fully captured by quality and/or efficiency shocks. Dealing with these different dimensions of heterogeneity is not only important in order to measure technical efficiency or TFP alone, but also from a welfare analysis point of view. For example, does an increase in market size push firms to charge higher or lower markups? Several recent theoretical papers have analyzed the relationship between market size, markups and welfare (Dhingra and Morrow, 2012 and Zhelobodko et al., 2012).

This paper’s contribution is to deal with these three issues in a more comprehensive way than the existing literature. Some existing studies ignore at least one or several of them while others address some of the issues but in a restrictive way. For example, Klette and Griliches (1996) allow for demand shocks but restrict to homogenous markups between firms. Other approaches, such as De Loecker and Warzynski (2012) and Dobbelaeere and Maïresse (2013) introduce heterogeneous markups but ignore the possibility of demand shocks. Papers such as Foster et al. (2008) allow for demand shocks but introduce the restriction that they must not be correlated with productivity shocks. We introduce a new framework for estimating the various firm-level dimensions of heterogeneity, namely demand, productivity and markups heterogeneity while leaving the correlation among the three unrestricted.

In our framework demand shocks shift demand in a way that is complementary to heterogeneity in firm-level productivity. However, there are at least two other important issues related to TFP estimation. The first one is that input quantity data at the firm-level is typically not available and input expenditure is used instead. The second is that many firms are actually multi-product and so the problem of how to assign inputs to outputs needs to be solved. In this paper we focus, for better comparability with previous studies, on single-product firms and do not deal with these two issues. Future research will expand in this direction. See De Loecker et al. (2014) for a joint treatment of input price bias and multi-product firms. See also Atalay (2014) for a quantification of the input price bias and Grieco et al. (2014) for a parsimonious methodology to deal with such a bias.

2 When only revenue data is available one can still identify markups as well as a composite of demand and TFP shocks. With firm-level output data one can in addition distinguish between TFP and demand shocks. See Martin (2014)
markups and that can be interpreted as a measure of quality of a firm’s products.

We subsequently use information on both the quantity and the value of Belgian manufacturing firm production over the period 1996-2007 to quantify productivity, markups and demand shocks. We show how these shocks are correlated among them, across time as well as with measures obtained from other approaches. We first document that demand shocks display at least as much variability across firms as productivity shocks. Productivity shocks are very strongly and negatively correlated with demand shocks in each of the four industries we consider. This finding is suggestive of a trade-off between the quality of a firm’s products and their production cost. Consider, for example, the car industry where there is the co-existence of manufacturers (like Nissan) producing many cars for a given amount of inputs (high productivity) and manufacturers (like Mercedes) producing less cars for a given amount of inputs (low productivity). To be more specific one of the most productive car plants in Europe is the Nissan factory located in Sunderland in the UK. In terms of sheer productivity measured as cars per employee it is nearly 100% more productive than a state of the art Mercedes plant near Rastatt in Germany. However, this hardly reflects a problem with the Mercedes plant. Rather, Mercedes and Nissan have very different demand shocks which leads to different prices as well different markups. Both plants are profitable and perhaps generate a very similar revenue productivity. Yet, their business model is quite different. They differentiate themselves in the quality-cost space and firms in the four industries we consider seem to do so too.

The second thing worth noting is that differences in markups across firms are reasonably well explained (in terms of $R^2$) by differences in demand shocks, productivity shocks and the capital stock. However, there remains a considerable amount of unexplained variation. Therefore, markups are far from being a residual dimension of heterogeneity in the data. Third, we find that variation in revenue TFP is actually attributable mainly to variation in demand across firms (demand shocks and markups) rather than in quantity TFP. Demand shocks typically explain more variation than markups.

We finally assess how and to what extent our three dimensions of heterogeneity allow to gain deeper and sharper insights on the understanding of two key firm outcomes: size and export status. We first start by showing that the usual positive correlation between revenue-based TFP and firm size, as measured by the number of employees, holds in our data. The availability of physical quantity data allows us to also look at the correlation between quantity-based TFP and firm size. We find such correlation to be positive, providing support to the mainstream theoretical framework based on differences across firms in term of their ability to turn inputs into output. Yet, within our framework we can go even further and ask whether and how demand heterogeneity also matters and how it interacts with heterogeneity in productivity. On the one hand we confirm that the positive correlation between quantity-based TFP and firm size is robust to the inclusion of demand shocks and markups heterogeneity. On the other hand, we find that demand heterogeneity is as important as productivity heterogeneity in understanding why some firm are larger than others. We also find that larger firms typically sell higher quality goods and charge lower markups. Finally, when considering firm export status, we show that the positive correlation between revenue-based TFP and firm export status commonly found in the trade literature equally holds in our data. It is also true with quantity-based TFP. When considering the other two dimensions, we find that the positive correlation between quantity-
based TFP and export status is robust to including demand shocks and markups heterogeneity. On the other hand, we find that demand heterogeneity is more important than productivity heterogeneity to draw the line between exporting and non-exporting firms. We also find that exporters typically sell higher quality goods and charge lower markups.

Our paper is related to the literature on firm TFP measurement on which Olley and Pakes (1996) has had a deep impact. The key endogeneity issue addressed in Olley and Pakes (1996) is omitted variables: the firm observes and takes decisions based on productivity shocks that are unobservable to the econometrician. Yet, the econometrician observes firm decisions (investments) that do not impact productivity today and that can (under certain conditions) be used as a proxy for productivity shocks. This proxy variable approach to tackle the issue of unobservable productivity shocks has been further developed in Levinsohn and Petrin (2003) and Ackerberg et al. (2006) and represents the current dominant framework.

In a recent paper De Loecker and Warzynski (2012) have extended this framework by explicitly allowing for another dimension of heterogeneity in the model, namely firm-specific markups, while providing an estimation strategy to separately identify productivity and markups. Building on Hall (1986) they show that, for a variety of market structures, there is a simple relationship between markups, the output elasticity of a variables input and the share of that input’s expenditure in total sales. This simple relationship allows them to readily compute firm-level markups from estimates of the parameters of the production function. In order to estimate the production function they build on Ackerberg et al. (2006).

However, this approach is consistent if the only unobserved driver of variations in prices - and thus markups - are productivity shocks. In order to improve upon De Loecker and Warzynski (2012) we need to make some explicit assumptions about the market structure. In our main specification we propose a generalized Dixit-Stiglitz monopolistic competition approach. It turns out that this does not only allow introducing separate demand and productivity shocks. In addition we can also allow for a possibly entirely independent third markup shock. In other words, our method can allow for markups that are either entirely determined by both productivity and/or demand shocks or are entirely independent from these other shocks or a combination of the previous two case. We further show in the Appendix that our approach can be generalised to consider more flexible functional forms than the one fitting the generalised Dixit Stiglitz framework.

Our interest in demand shocks is common to both De Loecker (2011) and Foster et al. (2008). De Loecker (2011) introduces demand shocks in a revenue-based production function model while relying on CES preferences and a common markup across varieties. This allows substituting for prices and get a tractable expression for firm revenue as a function of inputs, TFP and demand shocks. Compared to our framework, De Loecker (2011) does not allow for different markups across varieties. By contrast Foster et al. (2008) use data on both the quantity and the value of a firm’s production in order to disentangle productivity shocks from demand shocks. More specifically, they first recover production function coefficients from industry average cost shares and subsequently estimate a demand system featuring demand shocks measured as regression residuals and instrument firm price with firm firm TFP. Therefore, the identifying assumption allowing them to disentangle productivity...
shocks from demand shocks is that they are uncorrelated. In our framework we do not impose such an assumption. We instead draw on the additional structure we impose on the underlying firm behavior to disentangle the two shocks. In doing so we find them to be very strongly correlated with each other suggesting that, in our data, Foster et al. (2008) assumption of a zero correlation between productivity shocks and demand shocks is severely violated.

When comparing our measures with those obtained from other approaches, we find big differences in demand shocks but not much with respect to TFP and markups. As far as demand shocks are concerned there are two reasons why our shocks should be expected to be different from those of Foster et al. (2008). Foster et al. (2008) impose constant markups across firms and zero correlation between demand shocks and productivity shocks. We relax both assumptions and show that, though positively correlated (in between 0.2 and 0.4), the two sets of demand shocks are very different and can thus potentially lead to completely different conclusions when used to answer a specific research question. In terms of TFP we show that all standard quantity-based TFP measures (including ours) are very much correlated with each other. However, the measure developed here is based on a more comprehensive framework allowing unraveling many dimensions of heterogeneity potentially correlated with each other. These dimensions of heterogeneity, namely TFP shocks, demand shocks and markups, can in turn be used to get richer and deeper insights into important questions.

The rest of the paper is organized as follows. Section 2 provides our baseline econometric model and estimation procedure. We present our data in Section 3 while Section 4 contains estimation results as well as some descriptive statistics and correlations. We compare our measures with measures obtained from other approaches in Section 5 while in Section 6 we show how our three measures of heterogeneity can be used to get fresh insights into two key firm outcomes: size and export status. Section 7 concludes. Finally, in the Appendix we show how to extend our framework to more general preferences, production functions and processes for productivity and demand shocks.

2 Baseline model and estimation procedure

2.1 The Model

2.1.1 Production

Consider a Cobb-Douglas production technology with 3 production factors: labour (L), materials (M) and capital (K). Whereas labour and materials are perfectly flexible, capital is fixed in the short-run. We assume firms minimize costs and take the price of labour ($W_L$) and materials ($W_M$) as given.

Consequently, at any given point in time, each firm $i$ is dealing with the following short-run cost minimization problem:

\[ \text{minimize } C(L, M, K) \text{ subject to } P(L) = W_L, P(M) = W_M, P(K) = rK \]

We do not need to assume constant returns to scale ($\gamma=1$). It is also relatively straightforward to adapt the model to more general production technologies. See the Appendix for more details.

To simplify notation we do not use time indices unless needed to avoid ambiguity. We also ignore components that are constant across firms in a given time period as they will be controlled for by time dummies.
\[
\min_{L_i, M_i} \{L_i W_L + M_i W_M\} \quad \text{s.t.} \quad Q_i = A_i L_i^{\alpha_L} M_i^{\alpha_M} K_i^{\gamma - \alpha_M - \alpha_L}.
\]

where \(A_i\) is an idiosyncratic productivity shock observable to the firm but not the econometrician that we characterize in Section 2.2. First order conditions to this problem imply that:

\[
W_x = \chi_i \frac{Q_i}{X_{si}} \alpha_x
\]

for \(x \in \{L, M\}\) where \(X_{si}\) is either the amount of labor or of materials used by firm \(i\) and \(\chi_i\) is a Lagrange multiplier. In order to solve for \(\chi_i\) we can re-write (1) as:

\[
X_{si} = \chi_i \frac{Q_i}{W_x} \alpha_x
\]

and substitute \(X_{si}\) for \(M_i\) and \(L_i\) in the Cobb-Douglas expression:

\[
Q_i = \chi_i^{\alpha_L + \alpha_M} Q_i^{\alpha_L + \alpha_M} A_i \left( \frac{\alpha_L}{W_L} \right)^{\alpha_L} \left( \frac{\alpha_M}{W_M} \right)^{\alpha_M} K_i^{\gamma - \alpha_M - \alpha_L}.
\]

Hence:

\[
\chi_i = Q_i^{\alpha_L + \alpha_M - 1} \left( \frac{W_L}{\alpha_L} \right)^{\alpha_L} \left( \frac{W_M}{\alpha_M} \right)^{\alpha_M} K_i^{1 - \gamma / (\alpha_M + \alpha_L)}. \tag{2}
\]

From this formula we can derive the short-run cost function as:

\[
C_i = \chi_i \frac{Q_i}{W_L} \alpha_L W_L + \chi_i \frac{Q_i}{W_M} \alpha_M W_M = \chi_i Q_i (\alpha_L + \alpha_M)
\]

\[
= \left( \frac{Q_i}{A_i} \right)^{\frac{1}{\alpha_M}} \left( \frac{W_L}{\alpha_L} \right)^{\alpha_L} \left( \frac{W_M}{\alpha_M} \right)^{\alpha_M} K_i^{1 - \frac{\gamma}{\alpha_M + \alpha_L}} (\alpha_L + \alpha_M). \tag{3}
\]

Marginal cost thus satisfy the following property:

\[
\frac{\partial C_i}{\partial Q_i} = \frac{1}{\alpha_L + \alpha_M} \frac{C_i}{Q_i}.
\]

2.1.2 Demand and Market Structure

We consider a monopolistically competitive industry populated by a continuum of firms each producing one variety of a differentiated good. Each firm faces an idiosyncratic demand for its own variety and maximises profits while taking market aggregates as given. We further assume that firm demand can be described as:

\[
Q_i = D_i (\Lambda_i, P_i)
\]

and satisfies the property \(\frac{\partial \ln P_i}{\partial \ln \Lambda_i} = \frac{\partial \ln P_i}{\partial \ln Q_i} + 1\) where \(\Lambda_i\) is a firm-specific demand shock that is observable to the firm but not the econometrician, \(\frac{\partial \ln P_i}{\partial \ln Q_i} = -\frac{1}{\eta_i}\) and \(\eta_i\) is the elasticity of demand. We characterize
$\Lambda_i$ in Section 2.2 while in the Appendix we provide more insights on preferences satisfying the above property.

A simple but flexible case satisfying our property is the generalized CES preferences structure introduced by Spence (1976)\(^5\) that we adopt throughout this Section. In our baseline specification a representative consumer demand is thus obtained from the following problem:

$$\max_Q \left\{ \int_{i \in I} \frac{\eta_i}{\eta_i - 1} (\Lambda_i Q_i)^{\frac{\eta_i - 1}{\eta_i}} d_i \right\} \text{ s.t. } \int_i P_i Q_i d_i = B$$

where $B$ is the budget, $Q$ is a vector with elements $Q_i$ and the set of varieties is denoted by $I$. The first order condition to this problem implies:

$$P_i \kappa = \Lambda_i^{\frac{\eta_i - 1}{\eta_i}} Q_i^{\frac{1}{\eta_i}}$$

where $\kappa$ is a Lagrange multiplier. Re-arranging suggests that firm-level demand is:

$$Q_i = P_i^{-\eta_i} \Lambda_i^{\eta_i - 1} \kappa^{-\eta_i}.$$  

Profit maximization of firm $i$ then requires:

$$P_i = \mu_i \frac{\partial C_i}{\partial Q_i}$$

where the markup of firm $i$ is simply a function of the elasticity of demand: $\mu_i = \frac{\eta_i}{\eta_i - 1}$.

### 2.1.3 Some key properties

We derive here some useful properties that we will use in Section 2.2 to manipulate equations. First, from (1), (2), (3) and (5) we have:

$$\alpha_x = \frac{X_{xi} W_x}{X_i Q_i} = \frac{X_{xi} W_x}{C_i \alpha_L + \alpha_M} = \frac{X_{xi} W_x}{\mu_i P_i Q_i}.$$  

Therefore

$$\frac{\alpha_x}{\mu_i} = \frac{X_{xi} W_x}{P_i Q_i} = s_{xi}$$

where $s_{xi}$ is the expenditure share of factor $x \in \{L, M\}$ in firm $i$ revenue. This means that, for example, materials’ expenditure share is equal to the production function coefficient $\alpha_M$ divided by the markup $\mu_i$. This is a very useful property of our model delivering a simple way to measure markups:

$$\mu_i = \frac{\alpha_M}{s_{Mi}}.$$

---

\(^5\)See the Appendix for further details.
From (7) it is clear that the markup of a firm will be a scaling of the inverse of its materials’ expenditure share. Such share is typically observable in the data and does not require any estimation. However, we do need to estimate $\alpha_M$ in order to measure markups level.

The structure we impose also delivers useful properties for the revenue function. From now onwards we denote with small case the log of a variable (for example $\lambda_i$ denotes the natural logarithm of $\Lambda_i$).

Note that using $\mu_i = \frac{n_i}{n_{i-1}}$ as well as (4) we can write log revenue, up to an innocuous constant, as:

$$r_i = q_i + p_i = \frac{1}{\mu_i} (q_i + \lambda_i).$$  \hspace{1cm} (8)

(8) is a very important result in our model as it allows writing the revenue equation as a simple function of the markup as well as of the quantity and the demand shock. By substituting $q_i$ with the formula of the Cobb-Douglas we can transform this further as:

$$r_i = \alpha_L \mu_i (l_i - k_i) + \alpha_M \mu_i (m_i - k_i) + \gamma \mu_i k_i + \frac{1}{\mu_i} (a_i + \lambda_i).$$

Furthermore, by using (6) and (7), we finally get:

$$LHS_i \equiv \frac{r_i - s_{Li} (l_i - k_i) - s_{Mi} (m_i - k_i)}{s_{Mi}} = \frac{\gamma}{\alpha_M} k_i + \frac{1}{\alpha_M} (a_i + \lambda_i).$$  \hspace{1cm} (9)

The are two important features of (9) that will be better appreciated later on. First, the entire left-hand side ($LHS_i$) is made up of variables that are fully observable. Second, on the right hand side we have some key parameters ($\gamma$ and $\alpha_M$), log capital $k_i$ (which is given for a firm in the short-run) and two unobservable endogenous variables (that are known to the firm and drive its choices of inputs and pricing while being unobservable to the econometrician) entering linearly and with the same coefficient. By imposing enough structure to the process driving $a_i$ and $\lambda_i$, which we do next, (9) will allow us to estimate some key parameters.

### 2.2 Estimation Procedure

We now use the time index and assume, as it is typically done in models featuring unobservable productivity shocks, that $a_{it}$ evolves over time as stochastic Markov processes. We further assume that also $\lambda_{it}$ can be described by a Markov process:\footnote{For simplicity we assume that both processes are linear, i.e., together they are a VAR(1) process. We could, however, allow for more complex non-linear Markov processes as well as for firm fixed effects. See the Appendix for further details.}

$$a_{it} = \phi_a a_{i,t-1} + v_{ait}$$

$$\lambda_{it} = \phi_\lambda \lambda_{i,t-1} + v_{\lambda it}$$  \hspace{1cm} (10)

where $v_{a_{it}}$ and $v_{\lambda_{it}}$ can be correlated with each other. Before substituting (10) into (9) we need to find

\footnote{(8) holds as an equality in the generalized CES preferences structure we consider here. In other cases it holds as a local approximation. See the Appendix for further details.}
a convenient way to express \( a_{it-1} \) and \( \lambda_{it-1} \). By using (7) and (8) we have:

\[
\lambda_{it-1} = r_{it-1} \mu_{it-1} - q_{it-1} = r_{it-1} \frac{\alpha_M}{s_{Mit-1}} - q_{it-1}. \tag{11}
\]

At the same time plugging (11) into (9) and re-arranging yields:

\[
a_{it-1} = \alpha_M LHS_{it-1} - \gamma k_{it-1} - \left( r_{it-1} \frac{\alpha_M}{s_{Mit-1}} - q_{it-1} \right). \tag{12}
\]

Finally, by substituting (10) to (12) into (9) we obtain:

\[
LHS_{it} = \frac{\gamma}{\alpha_M} k_{it} + \phi_a LHS_{it-1} - \phi_a \frac{\gamma}{\alpha_M} k_{it-1}
+ \left( \phi_{\lambda} - \phi_{a} \right) \left( r_{it-1} \frac{1}{s_{Mit-1}} - \frac{1}{\alpha_M} q_{it-1} \right)
+ \frac{1}{\alpha_M} \left( \nu_{ait} + \nu_{\lambda it} \right). \tag{13}
\]

Equation (13) is key to us because it allows identifying two key parameters: \( \frac{\gamma}{\alpha_M} = \beta \) and \( \phi_a \). As will become clearer later on, it turns out that we do not actually need to estimate all of the model parameters to get measures of productivity shocks, demand shocks and markups. Indeed \( \beta, \phi_a \) and \( \gamma \) are sufficient. Using the revenue equation (13) we get estimates for \( \beta \) and \( \phi_a \). Using the quantity equation described below, along with estimates \( \hat{\beta} \) and \( \hat{\phi}_a \), we will in turn be able to estimate the returns to scale parameter \( \gamma \).

There are various way of estimating (13) and here we use perhaps the simplest one. More specifically, we rewrite (13) as the following linear regression:

\[
LHS_{it} = b_1 z_{1it} + b_2 z_{2it} + b_3 z_{3it} + b_4 z_{4it} + b_5 z_{5it} + u_{it} \tag{14}
\]

where \( z_{1it} = k_{it}, z_{2it} = LHS_{it-1}, z_{3it} = k_{it-1}, z_{4it} = r_{it-1} \frac{1}{s_{Mit-1}}, z_{5it} = q_{it-1}, u_{it} = \frac{1}{\alpha_M} \left( \nu_{ait} + \nu_{\lambda it} \right) \) as well as \( b_1 = \beta, b_2 = \phi_a, b_3 = -\phi_{\beta}, b_4 = (\phi_{\lambda} - \phi_{a}) \) and \( b_5 = -\left( \phi_{\lambda} - \phi_{a} \right) \frac{1}{\alpha_M} \). Given our assumptions the error term \( u_{it} \) in (14) is uncorrelated with all of the regressors. Therefore (14) can be estimated via simple OLS. After doing this we set \( \hat{\beta} = \hat{b}_1 \) and \( \hat{\phi}_a = \hat{b}_2 \) and do not exploit parameters’ constraints in the estimation.\(^8\)

We now turn to estimating \( \gamma \). Equation (6) implies \( \alpha_L = \mu_{it} s_{Lit} \) and \( \alpha_M = \mu_{it} s_{Mit} \). Firm log output \( q_{it} \) can thus be written as:

\[
q_{it} = \mu_{it} s_{Lit} (l_{it} - k_{it}) + \mu_{it} s_{Mit} (m_{it} - k_{it}) + \gamma k_{it} + a_{it}. \]

Further using (7) as well as the fact that \( \alpha_M = \frac{\gamma}{\beta} \) we get:

\(^8\)This means that, for example, we do not exploit the non-linear constraint \( b_3 = -1 \). We can certainly do this at the cost of using non-linear least squares. Furthermore, by exploiting parameters’ constraints we could actually also estimate \( \alpha_M \), and so \( \gamma \), from (14) without need for further estimations. However, identification of \( \alpha_M \) from (14) rests on the reduced form parameter \( (\phi_{\lambda} - \phi_{a}) \) being different from zero. In unreported results we generally fail to reject the hypothesis that \( (\phi_{\lambda} - \phi_{a}) \) is equal to zero.
\[ q_{it} = \frac{\gamma}{\beta} s_{L_{it}} (l_{it} - k_{it}) + \frac{\gamma}{\beta} (m_{it} - k_{it}) + \gamma k_{it} + a_{it} \]

where we replace \( \beta \) with \( \hat{\beta} \). Finally, using (10) to substitute for \( a_{it} \) and (12) to substitute for \( a_{it-1} \) we obtain:

\[
q_{it} = \frac{\gamma}{\hat{\beta}} s_{L_{it}} (l_{it} - k_{it}) + \frac{\gamma}{\hat{\beta}} (m_{it} - k_{it}) + \gamma k_{it} + \gamma k_{it} + \hat{\phi} a_{it} \gamma \hat{\beta} LHS_{it} - \hat{\phi} a_{it} k_{it} - \hat{\phi} a_{it} (r_{it-1} - \gamma \hat{\beta} S_{M_{it}} - q_{it-1}) + \nu_{ait}.
\]

(15)

Note that the only unobservable in (15) is the white noise term \( \nu_{ait} \) while the only parameter left unidentified is the scale parameter \( \gamma \). However, this time we cannot proceed with least squares because \( \nu_{ait} \) is correlated with the regressors and in particular with \( l_{it}, s_{L_{it}}, m_{it} \) and \( s_{M_{it}} \). Indeed, \( \nu_{ait} \) affects \( a_{it} \) and so input choices, pricing and revenues. Nonetheless we can, for example, identify \( \gamma \) from the following zero moment condition:

\[ E \{ \nu_{ait} k_{it} \} = 0 \]

applied to equation (15). We can implement this restriction in a linear regression framework by writing (15) as:

\[ LHS_{it} = b_6 z_{6it} + \nu_{ait} \]

(16)

where:

\[ LHS_{it} = q_{it} - \hat{\phi}_a q_{it-1} \]

\[
z_{6it} = \frac{1}{\hat{\beta}} s_{L_{it}} (l_{it} - k_{it}) + \frac{1}{\hat{\beta}} (m_{it} - k_{it}) + \frac{\hat{\phi}_a}{\hat{\beta}} LHS_{it-1} - \hat{\phi}_a k_{it-1} - \frac{\hat{\phi}_a}{\hat{\beta} S_{M_{it}} - q_{it-1}} + \hat{\phi}_a \]

as well as \( b_6 = \gamma \) and \( z_{6it} \) is instrumented with \( k_{it} \). We set \( \hat{\gamma} = b_6 \) and are in turn able to identify productivity shocks, demand shocks and markups:

\[
\hat{\alpha}_{it} = q_{it} - \frac{\hat{\gamma}}{\hat{\beta}} s_{L_{it}} (l_{it} - k_{it}) - \frac{\hat{\gamma}}{\hat{\beta}} (m_{it} - k_{it}) - \hat{\gamma} k_{it}
\]

\[
\hat{\mu}_{it} = \frac{\hat{\gamma}}{\hat{\beta} S_{M_{it}}}
\]

\[
\hat{\lambda}_{it} = \frac{\hat{\gamma}}{\hat{\beta} S_{M_{it}}} r_{it} - q_{it}
\]
Last but not least standard errors of $\hat{\beta}$, $\hat{\phi}$ and $\hat{\gamma}$ can be obtained via bootstrapping by re-sampling residuals in regressions (14) and (16).

3 Data

Our primary data consists of firm-level production data for Belgian manufacturing firms coming from the Prodcom database and provided by the National Bank of Belgium. Prodcom is a monthly survey of industrial production established by Eurostat for all EU countries in order to improve the comparability of production statistics across the EU by the use of a common product nomenclature called Prodcom (8-digit codes based on NACE 4-digits). Prodcom covers production of broad sectors C and D of NACE Rev. 1.1 (Mining and quarrying and manufacturing), except for sections 10 (Mining of coal and lignite), 11 (Extraction of crude petroleum and natural gas) and 23 (Manufacture of coke and refined petroleum products). Each firm with 20 employees or more or with a revenue greater than 3.5 million Euros in a given year has to fill the survey. Firms in the survey cover more than 90% of Belgian manufacturing production and the raw data is aggregated from the plant-level to the firm-level.

This gives us a sample of about 7,000 firms a year over the period of 1995 to 2009. Data is organised by product-year-month-firm. We use information on quantity (the unit of measurement depending on the specific product) and value (Euros) of production sold. We aggregate the data at the firm-year-product level. The same data has been previously used in Bernard et al. (2012b) in their analysis of carry along trade.

We also make use of more standard balance sheet data to get information on firms’ inputs. We build on annual firm accounts from the National Bank of Belgium. For this study, we selected those companies that filed a full-format or abbreviated balance sheet between 1996 and 2007 and with at least one full-time equivalent employee. The resulting dataset has been previously used in Behrens et al. (2013), Mion and Zhu (2013) and Muûls and Pisu (2009) and is representative of the Belgian economy. It includes information on FTE employment, material costs, capital stock and turnover. There are more than 15,000 manufacturing firms per year displaying non-missing values for these variables.

Last but not least we use standard EU-type micro trade data at the product-country-firm-month level over the period 1995-2008 provided by the National Bank of Belgium. From this data we simple borrow information on firm export status. The data has been previously used in Behrens et al. (2013), Mion and Zhu (2013) and Muûls (2015) among others. The three datasets are matched by the unique firm VAT identifier.

We focus on the period 1996-2007 for which all three datasets are available and during which there has not been any major change in data collection and data nomenclatures (such as NACE codes, Prodcom codes, etc.). We choose not to analyse multi-product firms in this paper and focus here on single-product firms.

As done in previous studies (Foster et al., 2008) estimations are run at a more aggregate level (labelled as ‘industry’ here), rather than at the finest available 8-digits product classification, in order to have
a sufficiently large number of observations. We use as industries NACE 3 digits-unit of measurement pairs (products belonging to a given NACE 3 digits code may have different units of measurement). This gives us about 200 industries to deal with.

The following cleaning and restrictions are applied such as to focus on single-product firms and have enough observations to both obtain precise estimates and be within a framework where monopolist competition can be considered as a reasonable approximation of the market structure:

- Consider only firm-year observations for which the value and quantity of production for all products (8-digit) are recorded
- Consider only firm-year observations for which employment, materials, sales and capital are available
- Aggregate production data at the 3 digit-unit of measurement level. See the Appendix for further details.
- Create for each firm-year the production value shares of its different 3 digit-unit products and keep a firm-year couple only if > 95% of production value is within a given 3 digit-unit: Single-product firms
- Apply small trimmings (1% up and down) based on capital intensity, share of intermediates in revenues and unit prices
- Consider only 3-digit sectors with more than 80 firms in each year

Doing so we end up with four industries on which we will focus our empirical analysis:

1. NACE 151: “Production, processing and preserving of meat and meat products”
2. NACE 212: “Manufacture of articles of paper and paperboard”
3. NACE 266: “Manufacture of articles of concrete, plaster and cement”
4. NACE 361: “Manufacture of furniture”

Figure 1 provides a visual image of the four industries as well as of some of the 8-digits products belonging to them. Two things are worth nothing. First, in order to estimate production functions at the industry level, we need to aggregate quantities produced of sometimes very different products. This is a problem also faced by other studies using physical production data\(^9\) and for which, to best of our knowledge, the current practice is to simply sum quantities across products within a firm. We improve on the existing practice by using the average (across firms) log price for each product \(j\) as a weight in the aggregation. We fully spell out in the Appendix the assumptions that make this approach meaningful. We further note that our results are qualitatively and, to a large extent also quantitatively, not affected by this choice.

\(^9\)See for example Foster et al. (2008)
The second thing worth nothing is that the industry with NACE 266 is not “Ready Mixed Concrete” but rather “Manufacture of articles of concrete, plaster and cement”. The former industry has been the object of numerous studies (Syverson, 2004; Foster et al., 2008) and is considered as a homogeneous good which is at best differentiated in terms of the geographic location of firms only. “Manufacture of articles of concrete, plaster and cement” is a rather different industry comprising products that, as depicted in Figure 1, are far from being homogeneous.

4 Results

In this Section we provide a number of descriptive statistics about our estimations, our measures of productivity shocks, demand shocks and markups and examine how the three dimensions of heterogeneity correlate with each other in a cross section as well as across time.

4.1 Descriptive statistics

Our main contribution is to provide a framework simultaneously allowing for heterogeneity in demand, TFP and markups. In this light it is desirable to first show how much heterogeneity coming from the demand side is present in the data. This is accomplished in Figure 2 where we show, for each of the four industries in our analysis, the plot of log price and log quantity stemming from the raw data. As one can notice firms sell very different quantities even though they charge the same price (different values on the X-axis for a given value of Y-axis). Differences are remarkably large considering that we are using log prices and quantities and provide direct evidence of the fact that firms do face heterogeneous demands, i.e., heterogeneous $\lambda_i$ and $\mu_i$ in our setting.

Turning to estimations, Table 1 show the mean and standard deviation of $\mu$, $a$ and $\lambda$ across firms in each of the four industries. Average markups range from 1.214 for “Manufacture of articles of concrete, plaster and cement” to 1.411 for “Manufacture of furniture”. Magnitudes are comparable to those obtained by De Loecker and Warzynski (2012) under different specifications. Though, the standard deviation is relatively small as compared to the about 0.5 reported in De Loecker and Warzynski (2012). This implies that the vast majority of our firm-level markups is indeed above the economically meaningful threshold of one. This is more evident in Figure 3 where we provide the density distribution of $\mu$.

As for productivity shocks and demand shocks the mean is not of much importance per se. The standard deviation is instead meaningful with the one of $a$ being considerably larger than the 0.26 reported in Foster et al. (2008) for physical TFP. Yet, it has to be considered that Foster et al. (2008) focus on industries characterized by rather homogeneous products for which is reasonable to expect less TFP variability across firms. As far as demand shocks are concerned out standard errors are instead in line with the 1.16 figure reported in Foster et al. (2008). Interestingly, in our analysis demand shocks display at least as much variability as productivity shocks. This can be further appreciated in Figure 4 where we provide the (centered) density distributions of $\lambda$ and $a$. 

13
Last but not least Table 1 provides our production function estimates for the coefficients of materials \((\alpha_M)\), labour \((\alpha_L)\) and scale \((\gamma)\) along with bootstrapped standard errors (200 replications). Estimates are in line with previous findings in the literature. In particular, there is evidence for moderate increasing returns to scale for quantity-based production functions which is in line with the findings of De Loecker (2011). Furthermore, bootstrapped standard errors suggest our estimates are overall rather precise.

4.2 Cross-sectional correlations

Table 2 reports cross-sectional correlations between \(\mu\), \(\lambda\) and \(a\) as well as log price \(p\). The Table provides several insights.

The first thing worth noting is that the correlation between demand and productivity shocks is far from being zero; with a zero correlation being the identification hypothesis for demand shocks in Foster et al. (2008). Productivity shocks \(a\) are very strongly and negatively correlated with demand shocks \(\lambda\) in each of the four industries we consider. This finding is suggestive of a trade-off between the quality of a firm’s products (as measured by \(\lambda\)) and their production cost (as measured by \(a\)). Consider, for example, the car industry where there is the co-existence of manufacturers (like Nissan) producing many cars for a given amount of inputs (high \(a\)) and manufacturers (like Mercedes) producing much less cars for a given amount of inputs (low \(a\)). At the same time, however, Mercedes produces cars of a higher quality in that, if the two cars were priced the same, Mercedes would sell much more cars (higher \(\lambda\)). Both manufacturers are profitable firms and perhaps have very similar revenue productivity (we will come back to this issue below). Yet, their business model is quite different. They differentiate themselves in the quality-cost space and so seem to do firms in our four industries.

The negative relationship between \(\lambda\) and \(a\) we find is far from being perfect, and so there are indeed firms in the data who have both high (low) \(\lambda\) and \(a\). Yet, the presence of a negative correlation between demand and productivity shocks is a first order feature of the data in our sample. This can be further appreciated in Figure 5 where we plot the (centered) values of \(\lambda\) and \(a\) in each of the four industries. The strength of the linear relationship is quite apparent from Figure 5. Furthermore, paying attention to scaling in the two axes, suggests a regression coefficient of -1, i.e., if the products of a firm are twice more valuable to consumers than the products of another firm (former firm has a \(\lambda\) twice as big as the latter)\(^{10}\) they will be (on average) twice more costly to produce (latter firm has an \(a\) twice as big as the former).\(^{11}\)

The second thing worth noting is that markups are reasonably well correlated with demand shocks. More specifically, we find that firms selling higher quality goods charge higher markups. The relationship between markups and \(a\) is instead much weaker and depends on the specific sector considered. Table 3 offers further insights on the relationship of markups with demand and productivity shocks. In a model in which the fundamental driver of heterogeneity in demand across firms is only \(\lambda\) (like

\(^{10}\) In the Appendix we show that, when starting from Direct Utility, \(\lambda\) has a quality interpretation coherent with this example.

\(^{11}\) This finding lends indirect support to the assumptions we make in the Appendix in order to aggregate quantities produced of different products within a firm.
in the Generalized Quadratic Utility case we consider in the Appendix) we would expect markups $\mu$ to vary across firms only to the extent that $\lambda$, $a$ and capital (with the latter two determining marginal costs) vary across firms.

In Table 3 we regress $\mu$ on $a$, $\lambda$ and capital $k$. Differences in markups across firms are reasonably well explained (in terms of $R^2$) by differences in demand shocks, productivity shocks and the capital stock. However, there is a considerable amount of unexplained heterogeneity. Therefore, the higher flexibility of the Generalised CES as compared to the Generalized Quadratic Utility seems to be needed in order to capture markups heterogeneity across firms. Furthermore, our results indicate, to the extent that a comparison can be made, that firms with higher productivity charge ceteris paribus, i.e., controlling for $\lambda$ and $k$, higher markups. This in line with preferences featuring increasing Relative Love for Variety (from which pro-competitive effects can be rationalised) and the presence of market distortions such that the market leads to too little selection with respect to the social optimum.\(^{12}\)

One last thing to note about Table 2 is the extremely strong correlation between log prices $p$ and TFP $a$ ranging from -0.916 to -0.956. This finding is in line with evidence reported in Foster et al. (2008) and was one of the grounds for their choice to instrument prices with TFP. Yet the correlation between prices and demand shocks is also very strong ranging from 0.691 to 0.926. The correlation with markups is instead much weaker ranging from not being significant to 0.213.

### 4.3 Correlations across time and predictive power

Numerous studies on productivity report a high degree of persistency across time while Foster et al. (2008) document a similar behavior for their measure of demand shocks. While based on a different approach and data type our analysis confirms these findings. Table 4 reports estimations. In each case we regress $a$, $\lambda$ and $\mu$ on their respective time lag. Both $a$ and $\lambda$ are characterized by a high degree of time persistency with autoregressive coefficients being around 0.9 and an $R^2$ of 0.8 or above. Markups are relatively less persistent with the autoregressive coefficient scoring around 0.8 and an $R^2$ of 0.7.

Before moving in the next Section to a more systematic comparison of our measures of heterogeneity with those obtained from other methodologies we end this Section providing a feeling of their predictive power. This is accomplished in Tables 5 to 7 where we regress, respectively, log price $p$, log quantity $q$ and log revenue $r$ on $a$, $\lambda$, $\mu$ and log capital $k$. In our model $p$, $q$ and $r$ should ultimately be a (non-linear) function of the primitives of the model: $a$, $\lambda$, $\mu$ and $k$. Though being an approximation the linear regression gives us, by means of the $R^2$, a feeling about the predictive power of the model. An inspection of the 3 Tables reveals our model scores extremely well for log prices with $R^2$ of around 0.95. As for log quantities results are more modest attaining $R^2$ of about 0.65. Considering the last Table we have $R^2$ slightly below those of $q$ which is not surprising given the results of the previous two Tables and the fact that $r=p+q$.

\(^{12}\)See Dhingra and Morrow (2012) and Zhelobodko et al. (2012) for further details.
5 Comparison to other methodologies

5.1 TFP

In order to gain insights into how and to what extent our methodology to measure TFP differs from other approaches we have computed additional TFP estimates based on:

- A GMM version of Olley and Pakes (1996): OP
- Industry costs shares as in Foster et al. (2008): FHS
- Ordinary Least Squares: OLS

For each case we have computed a revenue-based TFP (using revenue as a measure of output) and a quantity-based TFP (using physical quantity as a measure of output).

Our results suggest the following. First, there is considerable difference between revenue-based and quantity-based TFP. This has already been documented in Foster et al. (2008) using FHS productivity estimates. Our findings extend this results to a broader set of TFP measurement approaches while pointing to more substantial differences. For example, Foster et al. (2008) report a correlation between the two TFP of 0.64.\(^{13}\) We instead find the following correlations (across all industries while demeaning) between revenue-based and quantity-based TFP.\(^ {14}\)

- DLW, quantity and revenue based: 0.380***
- OP, quantity and revenue based: 0.0929***
- FHS, quantity and revenue based: 0.0863***
- OLS, quantity and revenue based: 0.0921***

\(* * * p<0.01, ** p<0.05, * p<0.1*

Second, the correlations (across all sectors while demeaning) between our quantity TFP measure \(a\) and quantity TFP measures computed with other methods are:

- DLW, quantity based: 0.866***
- OP, quantity based: 0.948***
- FHS, quantity based: 0.935***

\(^{13}\)The closest (to our) revenue TFP measure used in Foster et al. (2008) is what they label “Traditional TFP”.

\(^{14}\)As already noted earlier Foster et al. (2008) focus on industries characterized by rather homogeneous products for which is reasonable to expect less differences in prices and so a closer relationship between revenue-based and quantity-based TFP.
Therefore, the key message is that having quantity TFP is the key thing. The specific methodology is certainly important and our approach has the advantage of allowing TFP measurement within an framework where both markups and demand heterogeneity are present and potentially correlated with TFP shocks. However this is, to some extent, a second order problem. This lends support to Syverson (2011) in that: “the inherent variation in establishment- or firm-level microdata is typically so large as to swamp any small measurement-induced differences in productivity metrics.”

Where our framework delivers its full potential and ultimately provides a contribution is not in getting the TFP “more right” than in other methodologies but rather in allowing unraveling many dimensions of heterogeneity potentially correlated with each other. These dimensions of heterogeneity, namely TFP shocks, demand shocks and markups, can in turn be used to get richer and deeper insights into important questions. Consider, for example, revenue TFP. In the light of our framework revenue TFP will clearly be a mixture of physical productivity, markups and product quality but what we do not know is in which proportions.

Tables 8 to 11 answer this question. In these four Tables we regress DLW, OP, FHS and OLS revenue based productivities on $a \lambda$ and $\mu$ while reporting Beta coefficients. Interestingly, in most instances variation in revenue TFP is actually attributable mainly to variation in demand across firms ($\lambda$ and $\mu$) rather than in quantity TFP. At the same time $\lambda$ typically explains more variation than $\mu$ with the latter having sometimes a negative coefficient.

### 5.2 Markups

De Loecker and Warzynski (2012) provides the first fully-fledged framework to compute markups at the firm-level. We share with them a few assumptions and so the question of how our markups compare to theirs arises naturally.

In both frameworks markups are obtained as the ratio of the estimated output elasticity of a variable inputs free of adjustment costs (materials in our case with elasticity $\alpha_M$) to the share of that input’s expenditure in total sales. Therefore, provided both methods deliver the same estimate for $\alpha_M$, markups will be identical. Yet, even if the $\alpha_M$ are different the correlation between the two sets of markups will be one. Only to the extent that the getting the scale of markups right is important in the analysis they would thus be different. Our methodology has the advantage of not requiring proxies for markups while simultaneously allowing for demand shocks. DLW methodology has the advantage of not requiring both quantity and revenue data while being applicable outside the scope of monopolistic competition. We believe the right choice ultimately depends upon the specific data available and industry.

A substantial difference between the two methodologies arises when unobservable (to the firm) productivity shocks enter into the analysis. De Loecker and Warzynski (2012) propose a correction to the markups formula to take into account such shocks. When applying their correction in our data we get a (significant at the 1%) correlation (across all sectors while demeaning) between the two sets of

- OLS, quantity based: 0.948***
  *** p<0.01, ** p<0.05, * p<0.1
markups of only 0.0633. The difference is clearly substantial and calls for a serious evaluation of both the importance of productivity shocks unobservable to the firm as well as the capacity of the proxy variable approach to separate observable and unobservable (to the firm) shocks.

To gain further insights Table 12 provides average markups across the four industries computed with our methodology ($\mu$), DLW quantity-based productivity (DLW1) and DLW quantity-based productivity with correction for unobservable (to the firm) productivity shocks (DLW2).

5.3 Demand shocks

Foster et al. (2008) use US manufacturing firms production data containing information on both value and physical quantity to estimate quantity-based TFP as well as demand shocks. They measure demand shocks as the residual of a regression where log quantity is regressed on log price and the latter is instrumented with TFP obtained using industry costs shares to measure production function parameters (FHS TFP). The key identifying assumption in their framework is thus that productivity shocks are uncorrelated with demand shocks.

In light of our framework, Foster et al. (2008) approach is problematic for at least two reasons:

1. Markups are heterogeneous across firms: this means that the log price coefficient in their regression should be firm-specific. Within our framework we do not need to estimate those firm-specific coefficients because, based on our assumptions, they equal $-\eta_{it}$.

2. Demand shocks are strongly correlated with productivity shocks: this means that their IV strategy would not work in our data. Within our framework we do not need to take a stand on the correlation between demand and productivity shocks. (8) provides us with sufficient means to measure demand shocks once estimated TFP and markups.

In order to gain insights into the differences between the two approaches we have followed Foster et al. (2008) and computed demand shocks as the residual of a regression where log quantity is regressed on log price and the latter is instrumented with FHS TFP. Figure 6 shows a plot of $\lambda$ and FHS demand shocks for our four industries. Though positively correlated (correlations in Table 13 range from 0.231 to 0.414) the two sets of demand shocks are clearly quite different and can potentially lead to completely different conclusions when used to answer a specific research question.

To be fair, our $\lambda$ does not precisely correspond to the definition of demand shocks in Foster et al. (2008). Nevertheless, we can still define demand shocks as the residual component of model where log quantity is regressed over log price within our framework. From (4) we have $q_{it} = -\eta_{it} p_{it} + (\eta_{it} - 1) \lambda_{it} - \eta_{it} \ln \kappa_{it}$ and the residual component is thus $(\eta_{it} - 1) \lambda_{it} - \eta_{it} \ln \kappa_{it} = q_{it} + \eta_{it} p_{it}$. Figure 7 shows a plot of our residual demand shocks ($q_{it} + \eta_{it} p_{it}$) and FHS demand shocks. Again, though being

---

15 Foster et al. (2008) also control for a set of demand shifters, including a set of year dummies as well as the average income in the plant’s local market $m$ where local markets are defined based on Bureau of Economic Analysis’ Economic Areas. We also include in our regressions a full battery of year dummies. Yet, given the small size of Belgium we did not include any control for the plant’s local market income. Our IV estimations, available upon request, deliver highly (1%) significant coefficients for the log price coefficient in all four industries (point estimates are -1.4189, -1.4327, -.8850 and -1.1524 for industries 151, 212, 266 and 361 respectively).
positively correlated (correlations in Table 14 range from 0.058 to 0.299) the two sets of demand shocks are quite different.

In sum, we believe our methodology is to be preferred to measure demand shocks because it is actually more flexible and explicit than Foster et al. (2008) while having the same data requirements.

6 An application to firm size and export status

6.1 Firm size

Larger firms are typically found to be more productive than smaller ones under different measures of productivity. The empirical evidence is vast and encompasses both developed (Van Ark and Monnikhof, 1996) and developing (Van Biesebroeck, 2005) countries’ firms. At the same time there are several models consistent with this relationship being, on average, true in a cross section of firms (Jovanovic, 1982; Lucas Jr, 1978). Yet on the empirical side only revenue based measured of productivity have been used so far and, as seen in Section 5.1, these measures are a mixture of actual physical TFP, demand shocks and markups. This mirrors the situation on the theory side where the underlying differences across firms are in terms of their ability to turn inputs into output and not much in terms of the demand they face meaning that the distinction between revenue-based and quantity-based productivity is to a large extent immaterial. In what follows we build on the three measures of heterogeneity we construct to offer novel and shaper insights on the relationship between firm size, productivity and demand.

We first start by showing that the usual positive correlation between revenue-based TFP and firm size (as measured by the log number of employees) holds in our data. Tables 15 and 16 report results (Beta coefficients) of a linear estimation where the log number of employees of a firm is regressed on its OP and DLW revenue-based TFP respectively. Both sets of estimations convey the same message. Irrespective of the TFP measure used and industry we find a positive correlation between revenue-based TFP and firm size. The availability of physical quantity data allows to go one step further and look at the correlation between quantity-based TFP and firm size. This is done in Tables 17 and 18 reporting results (Beta coefficients) of a linear estimation where the log number of employees of a firm is regressed on its OP and DLW quantity-based TFP respectively. Tables 17 and 18 indicate that the positive correlation between firm size and revenue-based TFP extends to quantity-based TFP. This provides support to the mainstream theoretical framework based on differences across firms in term of their ability to turn inputs into output.

Yet, within our framework we can go even further and ask whether and how demand heterogeneity also matters and how it interacts with heterogeneity in productivity. This is achieved in Table 19 where we consider firm size regressed on $a$, $\lambda$ and $\mu$ while reporting Beta coefficients. On the one hand Table 19 confirms that the positive correlation between quantity-based TFP ($a$) and firm size is robust to the inclusion of demand shocks and markups heterogeneity. On the other hand, Table 19 broadens the spectrum of the analysis by pointing to the importance of demand heterogeneity in the understanding of why some firm are larger than others. Beta coefficients for $\lambda$ and $\mu$ are in general
smaller in magnitude than those of $a$. Yet, their combined effect is comparable to the one of quantity-based TFP while coefficient signs indicate that larger firms typically sell higher quality goods and charge lower markups.

### 6.2 Export status

Exporting firms are typically found to be more productive than non-exporters for different measures of productivity. The empirical evidence is vast (see Bernard et al., 2012a) while at the same time there are good reasons to believe that the direction of causality essentially goes from productivity to export status via a self-selection mechanism (Bernard and Jensen, 1999). This mechanism has been first fully spelled out in Melitz (2003) and has been the basis of a very prolific and influential theoretical and empirical literature (Helpman et al., 2015).

Yet, as in the firm size literature, empirical evidence is based only on revenue-based measures of productivity meaning that it is not clear whether the positive correlation stems from physical TFP and/or demand shocks and/or markups. On the theory side the mainstream approach relies on one dimension of heterogeneity across firms only (TFP) with heterogeneity in demand not receiving much attention.\(^{16}\) Yet, Melitz (2003) acknowledges that in his framework higher productivity can be either considered as producing a symmetric variety at lower marginal cost or producing a higher quality variety at equal cost. In what follows we build on the three measures of heterogeneity we construct to offer novel and sharper insights on the relationship between firm export status, productivity and demand.

We first start by showing that the usual positive correlation between revenue-based TFP and firm export status holds in our data. Tables 22 and 23 report results (Beta coefficients) of a linear estimation where the export status of a firm is regressed on its OP and DLW revenue-based TFP respectively. Both sets of estimations convey the same message. Irrespective of the TFP measure used and industry we find a positive correlation between revenue-based TFP and firm export status. The availability of physical quantity data allows to go one step further and look at the correlation between quantity-based TFP and export status. This is done in Tables 22 and 23 reporting results (Beta coefficients) of a linear estimation where firm export status is regressed on its OP and DLW quantity-based TFP respectively. Tables 22 and 23 indicate that the positive correlation between export status and revenue-based TFP extends to quantity-based TFP. This provides support to the mainstream theoretical framework based on differences across firms in term of their ability to produce at a lower marginal cost.

Yet, within our framework we can go even further and ask whether and how demand heterogeneity also matters and how it interacts with heterogeneity in productivity. This is achieved in Table 24 where we consider export status regressed on $a$, $\lambda$ and $\mu$ while reporting Beta coefficients. On the one hand Table 24 confirms that the positive correlation between quantity-based TFP ($a$) and firm export status is overall\(^{17}\) robust to the inclusion of demand shocks and markups heterogeneity. On the other hand, Table 19 expands the horizon of the analysis by pointing to the importance of demand

---

\(^{16}\)Some noticeable exceptions include Verhoogen (2008), Fajgelbaum et al. (2011) and Feenstra and Romalis (2014)

\(^{17}\)We fail to find a significant relationship in the Furniture industry. This might be due to the very high correlation (-0.910) between $a$ and $\lambda$ in this industry.
heterogeneity in the understanding of why some firm exports and others do not. Beta coefficients for \( \mu \) and especially for \( \lambda \) are in general larger in magnitude than those of \( a \). This suggest that demand heterogeneity is more important than differences in underlying physical productivity to draw the line between exporting and non-exporting firms. At the same time coefficient signs indicate that exporters typically sell higher quality goods and charge lower markups.

7 Conclusions

We provide a framework simultaneously allowing for heterogeneity in productivity, demand, and markups across firms while leaving the correlation among the three unrestricted. We are able to do this by imposing more explicit structure on preferences and firm behavior than in standard econometric models. We subsequently use information on both the quantity and the value of Belgian manufacturing firms production over the period 1996-2007 to quantify productivity, markups and demand shocks for four industries. We show how these shocks are correlated among them, across time as well as with measures obtained from other approaches. We finally assess how and to what extent our three dimensions of heterogeneity allow to gain deeper and sharper insights on the understanding of two key firm outcomes: export status and size.

There are many policy implications stemming from our analysis. For example, the capacity to distinguish the three components of revenue TFP is a crucial element for policy at different levels. At the micro level it makes a huge difference to know that some firms or industries lack in competitiveness because of poor physical TFP (due for example to low expenditure in process R&D) or poor product quality (due for example to low expenditure in product R&D). At the macro level one could look at aggregate revenue productivity cycles, like for example the severe downturn of UK aggregate revenue productivity since the financial crisis, not only in terms of changes in some underlying production capacity of the economy but also as changes in markups and/or consumers’ appreciation of firms’ products.

References


Table 1: Mean and Standard Deviation

<table>
<thead>
<tr>
<th>Industry</th>
<th>( \mu )</th>
<th>( \lambda )</th>
<th>( a )</th>
<th>( \alpha_M )</th>
<th>( \alpha_L )</th>
<th>( \gamma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>151</td>
<td>1.257</td>
<td>-13.230</td>
<td>14.000</td>
<td>0.976</td>
<td>0.170</td>
<td>1.109</td>
</tr>
<tr>
<td>212</td>
<td>1.304</td>
<td>-13.050</td>
<td>14.120</td>
<td>0.864</td>
<td>0.271</td>
<td>1.146</td>
</tr>
<tr>
<td>266</td>
<td>1.214</td>
<td>-13.190</td>
<td>14.160</td>
<td>0.813</td>
<td>0.238</td>
<td>1.068</td>
</tr>
<tr>
<td>361</td>
<td>1.411</td>
<td>-13.320</td>
<td>14.310</td>
<td>0.828</td>
<td>0.413</td>
<td>1.273</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard Deviation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>151</td>
<td>0.174</td>
<td>0.614</td>
<td>0.503</td>
<td>0.063*</td>
<td>0.011*</td>
<td>0.039*</td>
</tr>
<tr>
<td>212</td>
<td>0.229</td>
<td>0.768</td>
<td>0.591</td>
<td>0.048*</td>
<td>0.015*</td>
<td>0.032*</td>
</tr>
<tr>
<td>266</td>
<td>0.235</td>
<td>0.825</td>
<td>0.630</td>
<td>0.060*</td>
<td>0.018*</td>
<td>0.051*</td>
</tr>
<tr>
<td>361</td>
<td>0.313</td>
<td>1.163</td>
<td>1.160</td>
<td>0.033*</td>
<td>0.017*</td>
<td>0.036*</td>
</tr>
</tbody>
</table>

Notes: * indicates bootstrapped standard errors (200 replications).

Table 2: Correlations

<table>
<thead>
<tr>
<th></th>
<th>( \mu )</th>
<th>( \lambda )</th>
<th>( a )</th>
</tr>
</thead>
<tbody>
<tr>
<td>151</td>
<td>( \mu )</td>
<td>( \lambda )</td>
<td>( a )</td>
</tr>
<tr>
<td>( \mu )</td>
<td>1</td>
<td>( 0.187*** )</td>
<td>( -0.691*** )</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>( 0.417*** )</td>
<td>1</td>
<td>( -0.691*** )</td>
</tr>
<tr>
<td>( a )</td>
<td>( 0.742*** )</td>
<td>( -0.941*** )</td>
<td>1</td>
</tr>
<tr>
<td>( p )</td>
<td>0.020</td>
<td></td>
<td></td>
</tr>
<tr>
<td>212</td>
<td>( \mu )</td>
<td>( \lambda )</td>
<td>( a )</td>
</tr>
<tr>
<td>( \mu )</td>
<td>1</td>
<td>( 0.611*** )</td>
<td>( -0.767*** )</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>( 0.608*** )</td>
<td>1</td>
<td>( -0.910*** )</td>
</tr>
<tr>
<td>( a )</td>
<td>( -0.063 )</td>
<td>( -0.663*** )</td>
<td>1</td>
</tr>
<tr>
<td>( p )</td>
<td>0.213***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>266</td>
<td>( \mu )</td>
<td>( \lambda )</td>
<td>( a )</td>
</tr>
<tr>
<td>( \mu )</td>
<td>1</td>
<td>( 0.143*** )</td>
<td>( -0.956*** )</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>( 0.072*** )</td>
<td>1</td>
<td>( -0.910*** )</td>
</tr>
<tr>
<td>( a )</td>
<td>( -0.088*** )</td>
<td>( -0.910*** )</td>
<td>1</td>
</tr>
<tr>
<td>( p )</td>
<td>0.077***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>361</td>
<td>( \mu )</td>
<td>( \lambda )</td>
<td>( a )</td>
</tr>
<tr>
<td>( \mu )</td>
<td>1</td>
<td>( 0.791*** )</td>
<td>( -0.956*** )</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>( 0.926*** )</td>
<td>1</td>
<td>( -0.940*** )</td>
</tr>
<tr>
<td>( a )</td>
<td>( -0.088*** )</td>
<td>( -0.910*** )</td>
<td>1</td>
</tr>
<tr>
<td>( p )</td>
<td>0.077***</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: *** \( p < 0.01 \), ** \( p < 0.05 \), * \( p < 0.1 \).
Table 3: Regression of markup $\mu$ on $a$, $\lambda$ and log capital $k$

<table>
<thead>
<tr>
<th>Industry</th>
<th>151</th>
<th>212</th>
<th>266</th>
<th>361</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>.3369***</td>
<td>.2579***</td>
<td>.332***</td>
<td>.0155</td>
</tr>
<tr>
<td></td>
<td>(.015)</td>
<td>(.0314)</td>
<td>(.0141)</td>
<td>(.0356)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>.3259***</td>
<td>.3297***</td>
<td>.374***</td>
<td>.0369</td>
</tr>
<tr>
<td></td>
<td>(.0128)</td>
<td>(.014)</td>
<td>(.0089)</td>
<td>(.0355)</td>
</tr>
<tr>
<td>$k$</td>
<td>-.0368***</td>
<td>-.0447***</td>
<td>-.0259***</td>
<td>-.0801***</td>
</tr>
<tr>
<td></td>
<td>(.0044)</td>
<td>(.0055)</td>
<td>(.005)</td>
<td>(.006)</td>
</tr>
<tr>
<td># Obs</td>
<td>1235</td>
<td>770</td>
<td>1402</td>
<td>1566</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.6625</td>
<td>.6491</td>
<td>.6961</td>
<td>.1112</td>
</tr>
</tbody>
</table>

Notes: Time dummies are included in estimations but are not reported here. Bootstrapped standard errors in parenthesis (200 replications). *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 4: Regression of $a$, $\lambda$ and $\mu$ on their time lag

<table>
<thead>
<tr>
<th>Industry</th>
<th>151</th>
<th>212</th>
<th>266</th>
<th>361</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lag $a$</td>
<td>.9175***</td>
<td>.9477***</td>
<td>.9138***</td>
<td>.8372***</td>
</tr>
<tr>
<td></td>
<td>(.0224)</td>
<td>(.0211)</td>
<td>(.0172)</td>
<td>(.0202)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.8535</td>
<td>.8925</td>
<td>.8825</td>
<td>.7342</td>
</tr>
</tbody>
</table>

| Lag $\lambda$ | .8736*** | .8944*** | .9169*** | .8231*** |
|               | (.0238) | (.0246) | (.0204) | (.0212) |
| $R^2$        | .8135 | .8058 | .8396 | .7096 |

| Lag $\mu$   | .8013*** | .7949*** | .8493*** | .8743*** |
|             | (.0309) | (.0264) | (.019) | (.0225) |
| $R^2$       | .6869 | .7244 | .7381 | .7338 |

Notes: Time dummies are included in estimations but are not reported here. Bootstrapped standard errors in parenthesis (200 replications). *** $p<0.01$, ** $p<0.05$, * $p<0.1$. 
Table 5: Regression of log price $p$ on $a$, $\lambda$ and $\mu$ and log capital $k$

<table>
<thead>
<tr>
<th>Industry</th>
<th>151</th>
<th>212</th>
<th>266</th>
<th>361</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>-.8891***</td>
<td>-.6858***</td>
<td>-.6195***</td>
<td>-.4877***</td>
</tr>
<tr>
<td></td>
<td>(.0364)</td>
<td>(.0571)</td>
<td>(.0243)</td>
<td>(.0284)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>.0692*</td>
<td>.2005***</td>
<td>.286***</td>
<td>.48***</td>
</tr>
<tr>
<td></td>
<td>(.0346)</td>
<td>(.0508)</td>
<td>(.0235)</td>
<td>(.0268)</td>
</tr>
<tr>
<td>$\mu$</td>
<td>.4421***</td>
<td>-.0371</td>
<td>-.4334***</td>
<td>-.1729***</td>
</tr>
<tr>
<td></td>
<td>(.0881)</td>
<td>(.1275)</td>
<td>(.0564)</td>
<td>(.0513)</td>
</tr>
<tr>
<td>$k$</td>
<td>-.0729***</td>
<td>-.1229***</td>
<td>-.0674***</td>
<td>-.1351***</td>
</tr>
<tr>
<td></td>
<td>(.003)</td>
<td>(.0045)</td>
<td>(.0031)</td>
<td>(.0071)</td>
</tr>
</tbody>
</table>

# Obs 1235 770 1402 1566
$R^2$ .9496 .9357 .9415 .932

Notes: Time dummies are included in estimations but are not reported here. Bootstrapped standard errors in parenthesis (200 replications). *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 6: Regression of log quantity $q$ on $a$, $\lambda$ and $\mu$ and log capital $k$

<table>
<thead>
<tr>
<th>Industry</th>
<th>151</th>
<th>212</th>
<th>266</th>
<th>361</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>1.828***</td>
<td>1.074***</td>
<td>1.36***</td>
<td>.5458***</td>
</tr>
<tr>
<td></td>
<td>(.1563)</td>
<td>(.1975)</td>
<td>(.102)</td>
<td>(.0676)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>.9423***</td>
<td>.5549**</td>
<td>.7749***</td>
<td>-.3654***</td>
</tr>
<tr>
<td></td>
<td>(.1543)</td>
<td>(.1724)</td>
<td>(.1002)</td>
<td>(.0665)</td>
</tr>
<tr>
<td>$\mu$</td>
<td>-4.388***</td>
<td>-1.958***</td>
<td>-1.881***</td>
<td>-.7247***</td>
</tr>
<tr>
<td></td>
<td>(.4769)</td>
<td>(.4129)</td>
<td>(.2218)</td>
<td>(.0991)</td>
</tr>
<tr>
<td>$k$</td>
<td>.5814***</td>
<td>.7576***</td>
<td>.4271***</td>
<td>.6378***</td>
</tr>
<tr>
<td></td>
<td>(.0223)</td>
<td>(.029)</td>
<td>(.0198)</td>
<td>(.0217)</td>
</tr>
</tbody>
</table>

# Obs 1235 770 1402 1566
$R^2$ .6477 .7426 .5734 .6887

Notes: Time dummies are included in estimations but are not reported here. Bootstrapped standard errors in parenthesis (200 replications). *** $p<0.01$, ** $p<0.05$, * $p<0.1$. 

26
### Table 7: Regression of log revenue $r$ on $a$, $\lambda$ and $\mu$ and log capital $k$

<table>
<thead>
<tr>
<th>Industry</th>
<th>151</th>
<th>212</th>
<th>266</th>
<th>361</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>.9393***</td>
<td>.388**</td>
<td>.741***</td>
<td>.0581</td>
</tr>
<tr>
<td></td>
<td>(.1392)</td>
<td>(.1474)</td>
<td>(.0836)</td>
<td>(.0535)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>1.012***</td>
<td>.7554***</td>
<td>1.061***</td>
<td>.1146*</td>
</tr>
<tr>
<td></td>
<td>(.1353)</td>
<td>(.1332)</td>
<td>(.0803)</td>
<td>(.051)</td>
</tr>
<tr>
<td>$\mu$</td>
<td>-3.946***</td>
<td>-1.995***</td>
<td>-2.315***</td>
<td>-.8976***</td>
</tr>
<tr>
<td></td>
<td>(.4266)</td>
<td>(.3276)</td>
<td>(.1818)</td>
<td>(.0673)</td>
</tr>
<tr>
<td>$k$</td>
<td>.5085***</td>
<td>.6346***</td>
<td>.3596***</td>
<td>.5028***</td>
</tr>
<tr>
<td></td>
<td>(.0218)</td>
<td>(.022)</td>
<td>(.0189)</td>
<td>(.0198)</td>
</tr>
<tr>
<td># Obs</td>
<td>1235</td>
<td>770</td>
<td>1402</td>
<td>1566</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.6223</td>
<td>.733</td>
<td>.555</td>
<td>.5435</td>
</tr>
</tbody>
</table>

**Notes:** Time dummies are included in estimations but are not reported here. Bootstrap standard errors in parenthesis (200 replications). *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

### Table 8: DLW TFP revenue based regressed on $a$, $\lambda$ and $\mu$: BETA COEFFICIENTS

<table>
<thead>
<tr>
<th>Industry</th>
<th>151</th>
<th>212</th>
<th>266</th>
<th>361</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>1.236***</td>
<td>.9342***</td>
<td>.9488***</td>
<td>.9733***</td>
</tr>
<tr>
<td></td>
<td>(.0197)</td>
<td>(.0305)</td>
<td>(.0171)</td>
<td>(.0167)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>1.561***</td>
<td>1.448***</td>
<td>1.525***</td>
<td>1.063***</td>
</tr>
<tr>
<td></td>
<td>(.02)</td>
<td>(.0265)</td>
<td>(.0183)</td>
<td>(.0158)</td>
</tr>
<tr>
<td>$\mu$</td>
<td>-.6635***</td>
<td>-.485***</td>
<td>-.279***</td>
<td>-.2395***</td>
</tr>
<tr>
<td></td>
<td>(.0514)</td>
<td>(.063)</td>
<td>(.0427)</td>
<td>(.0226)</td>
</tr>
<tr>
<td># Obs</td>
<td>1233</td>
<td>769</td>
<td>1402</td>
<td>1561</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.558</td>
<td>.619</td>
<td>.639</td>
<td>.263</td>
</tr>
</tbody>
</table>

**Notes:** Time dummies are included in estimations but are not reported here. Bootstrap standard errors in parenthesis (200 replications). *** $p<0.01$, ** $p<0.05$, * $p<0.1$. 
Table 9: OP TFP revenue based regressed on $a$, $\lambda$ and $\mu$: BETA COEFFICIENTS

<table>
<thead>
<tr>
<th>Industry</th>
<th>151</th>
<th>212</th>
<th>266</th>
<th>361</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>.4062***</td>
<td>.3995***</td>
<td>.6407***</td>
<td>.6387***</td>
</tr>
<tr>
<td></td>
<td>(.011)</td>
<td>(.0124)</td>
<td>(.0101)</td>
<td>(.0079)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>.3678***</td>
<td>.461***</td>
<td>.821***</td>
<td>.6581***</td>
</tr>
<tr>
<td></td>
<td>(.011)</td>
<td>(.0111)</td>
<td>(.0095)</td>
<td>(.0077)</td>
</tr>
<tr>
<td>$\mu$</td>
<td>.4246***</td>
<td>.4022***</td>
<td>.1584***</td>
<td>.541***</td>
</tr>
<tr>
<td></td>
<td>(.0282)</td>
<td>(.0283)</td>
<td>(.0214)</td>
<td>(.0094)</td>
</tr>
<tr>
<td># Obs</td>
<td>1235</td>
<td>770</td>
<td>1402</td>
<td>1566</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.476</td>
<td>0.496</td>
<td>0.447</td>
<td>0.366</td>
</tr>
</tbody>
</table>

Notes: Time dummies are included in estimations but are not reported here. Bootstrapped standard errors in parenthesis (200 replications). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 10: FHS TFP revenue based regressed on $a$, $\lambda$ and $\mu$: BETA COEFFICIENTS

<table>
<thead>
<tr>
<th>Industry</th>
<th>151</th>
<th>212</th>
<th>266</th>
<th>361</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>.254**</td>
<td>.1858</td>
<td>.3797***</td>
<td>.2327**</td>
</tr>
<tr>
<td></td>
<td>(.0155)</td>
<td>(.0217)</td>
<td>(.0205)</td>
<td>(.0098)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>.2235*</td>
<td>.2986**</td>
<td>.5891***</td>
<td>.2656**</td>
</tr>
<tr>
<td></td>
<td>(.0156)</td>
<td>(.0184)</td>
<td>(.0194)</td>
<td>(.0097)</td>
</tr>
<tr>
<td>$\mu$</td>
<td>.2115**</td>
<td>.2716***</td>
<td>.0241</td>
<td>.4876***</td>
</tr>
<tr>
<td></td>
<td>(.044)</td>
<td>(.047)</td>
<td>(.0468)</td>
<td>(.0127)</td>
</tr>
<tr>
<td># Obs</td>
<td>1235</td>
<td>770</td>
<td>1402</td>
<td>1566</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.159</td>
<td>0.223</td>
<td>0.168</td>
<td>0.250</td>
</tr>
</tbody>
</table>

Notes: Time dummies are included in estimations but are not reported here. Bootstrapped standard errors in parenthesis (200 replications). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 11: OLS TFP revenue based regressed on $a$, $\lambda$ and $\mu$: BETA COEFFICIENTS

<table>
<thead>
<tr>
<th>Industry</th>
<th>151</th>
<th>212</th>
<th>266</th>
<th>361</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>.4153***</td>
<td>.3812***</td>
<td>.6245***</td>
<td>.6486***</td>
</tr>
<tr>
<td></td>
<td>(.0106)</td>
<td>(.0109)</td>
<td>(.0102)</td>
<td>(.0077)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>.3693***</td>
<td>.4339***</td>
<td>.783***</td>
<td>.6583***</td>
</tr>
<tr>
<td></td>
<td>(.0107)</td>
<td>(.0093)</td>
<td>(.0095)</td>
<td>(.0072)</td>
</tr>
<tr>
<td>$\mu$</td>
<td>.4255***</td>
<td>.4519***</td>
<td>.1813***</td>
<td>.572***</td>
</tr>
<tr>
<td></td>
<td>(.0284)</td>
<td>(.0234)</td>
<td>(.0232)</td>
<td>(.0085)</td>
</tr>
<tr>
<td># Obs</td>
<td>1235</td>
<td>770</td>
<td>1402</td>
<td>1566</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.464</td>
<td>0.523</td>
<td>0.429</td>
<td>0.392</td>
</tr>
</tbody>
</table>

Notes: Time dummies are included in estimations but are not reported here. Bootstrapped standard errors in parenthesis (200 replications). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. 

28
### Table 12: Comparison of average markups between our methodology and DLW

<table>
<thead>
<tr>
<th>Industry</th>
<th>( \mu )</th>
<th>markup DLW1</th>
<th>markup DLW2</th>
</tr>
</thead>
<tbody>
<tr>
<td>151</td>
<td>1.257</td>
<td>1.283</td>
<td>1.464</td>
</tr>
<tr>
<td>212</td>
<td>1.304</td>
<td>1.050</td>
<td>1.189</td>
</tr>
<tr>
<td>266</td>
<td>1.214</td>
<td>1.304</td>
<td>1.533</td>
</tr>
<tr>
<td>361</td>
<td>1.411</td>
<td>1.324</td>
<td>2.582</td>
</tr>
</tbody>
</table>

### Table 13: Correlation between \( \lambda \) and FHS demand shocks

<table>
<thead>
<tr>
<th>Industry</th>
<th>correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>151</td>
<td>0.294***</td>
</tr>
<tr>
<td>212</td>
<td>0.414***</td>
</tr>
<tr>
<td>266</td>
<td>0.238***</td>
</tr>
<tr>
<td>361</td>
<td>0.231***</td>
</tr>
</tbody>
</table>

*Notes: *** \( p < 0.01 \), ** \( p < 0.05 \), * \( p < 0.1 \).*

### Table 14: Correlation between our residual demand shocks and FHS demand shocks

<table>
<thead>
<tr>
<th>Industry</th>
<th>correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>151</td>
<td>0.217***</td>
</tr>
<tr>
<td>212</td>
<td>0.299***</td>
</tr>
<tr>
<td>266</td>
<td>0.058</td>
</tr>
<tr>
<td>361</td>
<td>0.162***</td>
</tr>
</tbody>
</table>

*Notes: *** \( p < 0.01 \), ** \( p < 0.05 \), * \( p < 0.1 \).*
Table 15: Log number of employees regressed on OP revenue-based TFP: BETA COEFFICIENTS

<table>
<thead>
<tr>
<th>Industry</th>
<th>OP TFP Revenue</th>
<th># Obs</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>151</td>
<td>0.5882***</td>
<td>1235</td>
<td>0.348</td>
</tr>
<tr>
<td>212</td>
<td>0.7237***</td>
<td>770</td>
<td>0.524</td>
</tr>
<tr>
<td>266</td>
<td>0.557***</td>
<td>1402</td>
<td>0.316</td>
</tr>
<tr>
<td>361</td>
<td>0.6863***</td>
<td>1566</td>
<td>0.470</td>
</tr>
</tbody>
</table>

Notes: Time dummies are included in estimations but are not reported here. Bootstrapped standard errors in parenthesis (200 replications). *** p<0.01, ** p<0.05, * p<0.1.

Table 16: Log number of employees regressed on DLW revenue-based TFP: BETA COEFFICIENTS

<table>
<thead>
<tr>
<th>Industry</th>
<th>DLW TFP revenue</th>
<th># Obs</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>151</td>
<td>0.7855***</td>
<td>1233</td>
<td>0.620</td>
</tr>
<tr>
<td>212</td>
<td>0.7664***</td>
<td>769</td>
<td>0.590</td>
</tr>
<tr>
<td>266</td>
<td>0.7994***</td>
<td>1402</td>
<td>0.640</td>
</tr>
<tr>
<td>361</td>
<td>0.8514***</td>
<td>1561</td>
<td>0.719</td>
</tr>
</tbody>
</table>

Notes: Time dummies are included in estimations but are not reported here. Bootstrapped standard errors in parenthesis (200 replications). *** p<0.01, ** p<0.05, * p<0.1.

Table 17: Log number of employees regressed on OP quantity-based TFP: BETA COEFFICIENTS

<table>
<thead>
<tr>
<th>Industry</th>
<th>OP TFP quantity</th>
<th># Obs</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>151</td>
<td>0.353***</td>
<td>1235</td>
<td>0.131</td>
</tr>
<tr>
<td>212</td>
<td>0.3565***</td>
<td>770</td>
<td>0.135</td>
</tr>
<tr>
<td>266</td>
<td>0.2492***</td>
<td>1402</td>
<td>0.0751</td>
</tr>
<tr>
<td>361</td>
<td>0.212***</td>
<td>1566</td>
<td>0.0461</td>
</tr>
</tbody>
</table>

Notes: Time dummies are included in estimations but are not reported here. Bootstrapped standard errors in parenthesis (200 replications). *** p<0.01, ** p<0.05, * p<0.1.

Table 18: Log number of employees regressed on DLW quantity-based TFP: BETA COEFFICIENTS

<table>
<thead>
<tr>
<th>Industry</th>
<th>DLW TFP quantity</th>
<th># Obs</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>151</td>
<td>0.3285***</td>
<td>1233</td>
<td>0.113</td>
</tr>
<tr>
<td>212</td>
<td>0.3737***</td>
<td>769</td>
<td>0.148</td>
</tr>
<tr>
<td>266</td>
<td>0.1384***</td>
<td>1402</td>
<td>0.0329</td>
</tr>
<tr>
<td>361</td>
<td>0.3854***</td>
<td>1561</td>
<td>0.149</td>
</tr>
</tbody>
</table>

Notes: Time dummies are included in estimations but are not reported here. Bootstrapped standard errors in parenthesis (200 replications). *** p<0.01, ** p<0.05, * p<0.1.
Table 19: Log number of employees regressed on $a$, $\lambda$, and $\mu$: BETA COEFFICIENTS

<table>
<thead>
<tr>
<th>Industry</th>
<th>151</th>
<th>212</th>
<th>266</th>
<th>361</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>1.11***</td>
<td>1.021***</td>
<td>.9931***</td>
<td>.341***</td>
</tr>
<tr>
<td></td>
<td>(.1236)</td>
<td>(.1523)</td>
<td>(.0764)</td>
<td>(.0427)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>.7612***</td>
<td>.4728***</td>
<td>.4961***</td>
<td>.2496***</td>
</tr>
<tr>
<td></td>
<td>(.1339)</td>
<td>(.1643)</td>
<td>(.0752)</td>
<td>(.0439)</td>
</tr>
<tr>
<td>$\mu$</td>
<td>-.5233***</td>
<td>-.5128***</td>
<td>-.2275***</td>
<td>-.2184***</td>
</tr>
<tr>
<td></td>
<td>(.275)</td>
<td>(.3911)</td>
<td>(.1655)</td>
<td>(.0636)</td>
</tr>
</tbody>
</table>

# Obs 1235 770 1402 1566
$R^2$ 0.291 0.290 0.289 0.0708

Notes: Time dummies are included in estimations but are not reported here. Bootstrapped standard errors in parenthesis (200 replications). *** p<0.01, ** p<0.05, * p<0.1.

Table 20: Export status regressed on OP revenue-based TFP: BETA COEFFICIENTS

<table>
<thead>
<tr>
<th>Industry</th>
<th>151</th>
<th>212</th>
<th>266</th>
<th>361</th>
</tr>
</thead>
<tbody>
<tr>
<td>OP TFP Revenue</td>
<td>.4428***</td>
<td>.3651***</td>
<td>.2075***</td>
<td>.4219***</td>
</tr>
<tr>
<td></td>
<td>(.0128)</td>
<td>(.0202)</td>
<td>(.0232)</td>
<td>(.0153)</td>
</tr>
</tbody>
</table>

# Obs 1235 770 1402 1566
$R^2$ 0.223 0.161 0.0512 0.206

Notes: Time dummies are included in estimations but are not reported here. Bootstrapped standard errors in parenthesis (200 replications). *** p<0.01, ** p<0.05, * p<0.1.

Table 21: Export status regressed on DLW revenue-based TFP: BETA COEFFICIENTS

<table>
<thead>
<tr>
<th>Industry</th>
<th>151</th>
<th>212</th>
<th>266</th>
<th>361</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLW TFP revenue</td>
<td>.4249***</td>
<td>.4094***</td>
<td>.4072***</td>
<td>.4378***</td>
</tr>
<tr>
<td></td>
<td>(.0705)</td>
<td>(.0603)</td>
<td>(.0533)</td>
<td>(.0319)</td>
</tr>
</tbody>
</table>

# Obs 1233 769 1402 1561
$R^2$ 0.209 0.196 0.172 0.219

Notes: Time dummies are included in estimations but are not reported here. Bootstrapped standard errors in parenthesis (200 replications). *** p<0.01, ** p<0.05, * p<0.1.
Table 22: Export status regressed on OP quantity-based TFP: BETA COEFFICIENTS

<table>
<thead>
<tr>
<th>Industry</th>
<th>151</th>
<th>212</th>
<th>266</th>
<th>361</th>
</tr>
</thead>
<tbody>
<tr>
<td>OP TFP quantity</td>
<td>.3646***</td>
<td>.2247***</td>
<td>.0359</td>
<td>.2706***</td>
</tr>
<tr>
<td></td>
<td>(.0134)</td>
<td>(.0172)</td>
<td>(.0164)</td>
<td>(.0079)</td>
</tr>
<tr>
<td># Obs</td>
<td>1235</td>
<td>770</td>
<td>1402</td>
<td>1566</td>
</tr>
<tr>
<td>R²</td>
<td>0.161</td>
<td>0.0804</td>
<td>0.0105</td>
<td>0.102</td>
</tr>
</tbody>
</table>

Notes: Time dummies are included in estimations but are not reported here. Bootstrapped standard errors in parenthesis (200 replications). *** p<0.01, ** p<0.05, * p<0.1.

Table 23: Export status regressed on DLW quantity-based TFP: BETA COEFFICIENTS

<table>
<thead>
<tr>
<th>Industry</th>
<th>151</th>
<th>212</th>
<th>266</th>
<th>361</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLW TFP quantity</td>
<td>.2649***</td>
<td>.2572***</td>
<td>.0665*</td>
<td>.3248***</td>
</tr>
<tr>
<td></td>
<td>(.0256)</td>
<td>(.0243)</td>
<td>(.0215)</td>
<td>(.0083)</td>
</tr>
<tr>
<td># Obs</td>
<td>1233</td>
<td>769</td>
<td>1402</td>
<td>1561</td>
</tr>
<tr>
<td>R²</td>
<td>0.0995</td>
<td>0.0961</td>
<td>0.0136</td>
<td>0.134</td>
</tr>
</tbody>
</table>

Notes: Time dummies are included in estimations but are not reported here. Bootstrapped standard errors in parenthesis (200 replications). *** p<0.01, ** p<0.05, * p<0.1.

Table 24: Export status regressed on a, λ and µ: BETA COEFFICIENTS

<table>
<thead>
<tr>
<th>Industry</th>
<th>151</th>
<th>212</th>
<th>266</th>
<th>361</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>.5184***</td>
<td>.3714***</td>
<td>.4297***</td>
<td>.108</td>
</tr>
<tr>
<td></td>
<td>(.0523)</td>
<td>(.0388)</td>
<td>(.0407)</td>
<td>(.0243)</td>
</tr>
<tr>
<td>λ</td>
<td>.6062***</td>
<td>.6067***</td>
<td>.7451***</td>
<td>.0205</td>
</tr>
<tr>
<td></td>
<td>(.0481)</td>
<td>(.0364)</td>
<td>(.0402)</td>
<td>(.0244)</td>
</tr>
<tr>
<td>µ</td>
<td>-.4762***</td>
<td>-.3782***</td>
<td>-.2456***</td>
<td>-.221***</td>
</tr>
<tr>
<td></td>
<td>(.1321)</td>
<td>(.1027)</td>
<td>(.0889)</td>
<td>(.0393)</td>
</tr>
<tr>
<td># Obs</td>
<td>1235</td>
<td>770</td>
<td>1402</td>
<td>1566</td>
</tr>
<tr>
<td>R²</td>
<td>0.126</td>
<td>0.119</td>
<td>0.117</td>
<td>0.0889</td>
</tr>
</tbody>
</table>

Notes: Time dummies are included in estimations but are not reported here. Bootstrapped standard errors in parenthesis (200 replications). *** p<0.01, ** p<0.05, * p<0.1.
Figure 1: Industries and Products in our sample

**Sector 151: Meat and Meat Products**
- code 15131215: Sausages not of liver
- code 15131225: Preparations of animal liver (incl. pates & pastes other than in sausage) food preparations containing >20% of meat (excl. sausages/homogenized preparations, of goose or duck)
- code 15131259: Preparations of pork (incl. mixtures; fats of any kind or origin, excl. prepared dishes, sausages and similar products, pates and pastes, homogenized preparations)

**Sector 212: Articles of Paper**
- code 21231230: Envelopes of paper or paperboard
- code 21241190: Wallpaper and other wall coverings; window transparencies of paper, n.e.c.

**Sector 266: Articles of Concrete**
- Code 26611130: Building blocks and bricks of cement; concrete or artificial stone
- Code 26611200: Prefabricated structural components for building, of cement

**Sector 361: Furniture**
- Code 36111250: Upholstered seats with wooden frames (incl. three piece suites) (excl. swivel seats)
- Code 36111290: Non-upholstered seats with wooden frames (excl. swivel seats)
- Code 36111250: Upholstered seats with wooden frames (incl. three piece suites) (excl. swivel seats)
Figure 2: How Important is Demand Heterogeneity? Plot of Log Price and Log Quantity

Plot of Log Price on Log Quantity

Meat

Paper

Concrete

Furniture
Figure 3: Density of $\mu$

Meat

Paper

Concrete

Furniture

kernel = epanechnikov, bandwidth = 0.0321

kernel = epanechnikov, bandwidth = 0.0520

kernel = epanechnikov, bandwidth = 0.0479

kernel = epanechnikov, bandwidth = 0.0517
Figure 4: Density of $\lambda$ and $a$

Meat

Paper

Concrete

Furniture

- $\lambda$ and $a$ are plotted for different materials: Meat, Paper, Concrete, and Furniture.
- The x-axis represents $x$, and the y-axis represents the density.
- Each graph compares the density of $\lambda$ (blue line) and $a$ (red line).

36
Figure 5: Plot of $\lambda$ and $a$
Figure 6: Plot of $\lambda$ and FHS demand shocks

Plot of Lambda on FHS demand shock

Meat

Paper

Concrete

Furniture
Figure 7: Plot of our residual demand shocks and FHS demand shocks

Plot of Residual demand shock on FHS demand shock

Meat

Concrete

Paper

Furniture
Appendix

In Appendices A to C we show how to extend our framework to more general preferences, production functions and processes for productivity and demand shocks. In Appendix D we instead spell out the assumptions allowing us to deal with the issue of aggregation.

A Preferences

There are various ways to extend our model and estimation methodology to preferences other than the baseline generalized CES case. One possibility is to identify a class of preferences under which the key equation (8) holds as a first-order linear approximation and modify the estimation procedure accordingly. Another possibility is to fully specify an alternative preference structure and work out the corresponding algebra for the estimation equations. We discuss the former case in Section A.1 while the latter case is presented in Section A.2

A.1 First-order linear approximation

The key property we want preferences to satisfy is that the elasticity of prices with respect to output quantity differs from the elasticity of prices with respect to the demand shock by one: \( \frac{\partial p_i}{\partial \lambda_i} = \frac{\partial p_i}{\partial q_i} + 1 \). There are different ways of achieving this. One way is to start from direct utility. Suppose a representative consumer maximises a differentiable utility function \( U(\cdot) \) subject to budget \( B \):

\[
\max_{\tilde{Q}} \left\{ U(\tilde{Q}) \right\} \text{ s.t. } \int P_i Q_i di - B = 0
\]

where \( \tilde{Q} \) is a vector of elements \( \Lambda_i Q_i \). Therefore, while the representative consumer chooses quantities \( Q \), these quantities enter into the utility function as \( \tilde{Q} \) and \( \Lambda_i \) can be interpreted as a measure of quality for variety \( i \). For example, in the symmetric (with respect to \( \tilde{Q} \)) varieties case, the representative consumer would be indifferent between having one more unit of a variety with \( \Lambda_i = \bar{\Lambda} \) or \( \bar{\Lambda} \) more units of a a variety with \( \Lambda_i = 1 \).

The first order conditions of the utility maximization problem imply:

\[
\frac{\partial U}{\partial \tilde{Q}_i} = \frac{\partial U}{\partial Q_i} \frac{\partial \tilde{Q}_i}{\partial Q_i} = \frac{\partial U}{\partial Q_i} \Lambda_i = \kappa P_i
\]

where \( \kappa \) is a Lagrange multiplier and \( \frac{\partial \tilde{Q}_i}{\partial Q_i} = \Lambda_i \). Taking logs we have:

\[
\ln \frac{\partial U}{\partial \tilde{Q}_i} + \lambda_i = \ln \kappa + p_i.
\]

Solving all of these conditions would give us demand functions for all varieties including that of firm \( i \). However, even if we knew the exact form of \( U(\cdot) \), this might be tricky to work out. Nonetheless,
(18) already tells us a lot about the shape of such demand functions. On the one hand, differentiating both sides with respect to \( q_i \) yields:

\[
\frac{\partial p_i}{\partial q_i} = \frac{\partial \ln \frac{\partial U}{\partial \tilde{Q}_i}}{\partial \tilde{q}_i} = \frac{\partial \ln \frac{\partial U}{\partial \tilde{Q}_i}}{\partial \tilde{q}_i} = \frac{\partial \ln \frac{\partial U}{\partial \tilde{Q}_i}}{\partial \tilde{q}_i},
\]

(19)

where \( \frac{\partial \tilde{q}_i}{\partial q_i} = 1 \) and \( \frac{\partial p_i}{\partial q_i} \equiv -\frac{1}{\eta_i} \). On the other hand, keeping in mind that \( \frac{\partial \tilde{q}_i}{\partial \lambda_i} = 1 \) differentiation of both sides with respect to \( \lambda_i \) gives:

\[
\frac{\partial p_i}{\partial \lambda_i} = \frac{\partial \ln \frac{\partial U}{\partial \tilde{Q}_i}}{\partial \lambda_i} + 1 = \frac{\partial \ln \frac{\partial U}{\partial \tilde{Q}_i}}{\partial \tilde{q}_i} + 1 = \frac{\partial \ln \frac{\partial U}{\partial \tilde{Q}_i}}{\partial \tilde{q}_i} + 1 = 1 - \frac{1}{\eta_i}
\]

i.e., the elasticity of the price with respect to quantity differs from the elasticity of the price with respect to the demand shock by one. This is the key property needed in our framework.

Let us now consider the implications of these results. Using (18) we can write log revenue \( r_i \) (up to a constant) as:

\[
r_i = p_i + q_i = \ln \frac{\partial U}{\partial \tilde{Q}_i} + \lambda_i + q_i = \ln \frac{\partial U}{\partial \tilde{Q}_i} + \tilde{q}_i.
\]

Differentiating both sides with respect to \( \tilde{q}_i \) and making use of (19) we have:

\[
\frac{\partial r_i}{\partial \tilde{q}_i} = \frac{\partial \ln \frac{\partial U}{\partial \tilde{Q}_i}}{\partial \tilde{q}_i} + 1 = \frac{1}{\eta_i} + 1 = \frac{1}{\mu_i}
\]

and so finally get:

\[
\Delta r_i \approx \frac{1}{\mu_i} \Delta \tilde{q}_i = \frac{1}{\mu_i} \Delta (q_i + \tilde{q}_i).
\]

(20)

Therefore, for any preferences structure that can be used to model monopolistic competition and that can be conveniently described by a direct utility we can, starting from the baseline formulation \( U(Q) \), introduce quality in such a way that (20) is satisfied. The advantage of (20) is that it can be directly used for estimations without need to explicitly solve for the demand functions of the different varieties.

One interesting example is the Generalized CES (Spence, 1976):

\[
U(\tilde{Q}) = \int_{i \in I} a_i (\tilde{Q}_i)^{b_i} di = \int_{i \in I} a_i \Lambda_i^{b_i} (Q_i)^{b_i} di
\]

where \( b_i = 1 - \frac{1}{\eta_i} \). If we further impose \( a_i = \frac{\eta_i}{\eta_i - 1} \) not only (20) holds but we actually get (8): \( r_i = \frac{1}{\mu_i} (q_i + \lambda_i) \). This is our benchmark case.

Another way of getting the key property needed is to start from demand functions and work out some constraints. Consider, for example, the Generalised Quadratic Utility (Di Comite et al., 2014):
\[ U(Q) = \int_{i \in I} a_i Q(i) \, di - \frac{1}{2} \int_{i \in I} b_i [Q(i)]^2 \, di - \frac{c_i}{2} \left[ \int_{i \in I} Q(i) \, di \right]^2 + Q_0 \]

where \( Q_0 \) is a numéraire good. Because of the presence of a numéraire good the Generalised Quadratic Utility does not fit the above described framework. Yet, one can easily derive the inverse demand function in this case and impose constraints on \( a_i, b_i \) and \( c_i \) such that \( \frac{\partial p_i}{\partial \alpha_i} = \frac{\partial p_i}{\partial q_i} + 1 \). This is obtained by setting \( a_i = a_1 \Lambda_i, b_i = b_1 (\Lambda_i)^2 \) and \( c_i = c_1 \Lambda_i \), where \( a_1, b_1 \) and \( c_1 \) are positive constants. The inverse demand will thus be:

\[ P_i = d_1 \Lambda_i - b_1 \Lambda_i^2 Q_i, \]

where \( d_1 = (a_1 - c_1 \bar{Q}) \) and \( \bar{Q} = \int_{i \in I} Q(i) \). The demand shock enters the demand function in order to ensure that \( \frac{\partial p_i}{\partial \alpha_i} = \frac{-b_1 Q_i \Lambda_i^2}{P_i} + 1 = \frac{\partial p_i}{\partial q_i} + 1 \). Therefore:

\[ \frac{\partial r_i}{\partial q_i} = \frac{\partial q_i}{\partial q_i} + \frac{\partial p_i}{\partial q_i} = 1 + \frac{\partial p_i}{\partial \lambda_i} = \frac{1}{\mu_i} = \frac{\partial q_i}{\partial \lambda_i} + \frac{\partial p_i}{\partial \lambda_i} = \frac{\partial r_i}{\partial \lambda_i} \]

and we obtain from this (20):

\[ \Delta r_i \approx \frac{1}{\mu_i} (\Delta q_i + \Delta \lambda_i) = \frac{1}{\mu_i} \Delta (q_i + \lambda_i) = \frac{1}{\mu_i} \Delta q_i. \]

In the case of Translog preferences, which are described using the expenditure function rather than the direct utility, one can get the desired property it in a similar manner.

Moving to the estimation strategy one needs to decide which type of local approximation (\( \Delta \)) to use. One possibility is consider differences in log revenue (as well as in other variables needed in the estimation) between a reference firm \( r \) and any other firm \( i \). This approach has the advantage of allowing measuring demand, productivity ad markups for all firms as deviations with respect to firm \( r \). However, the drawback is that for firms that are very different from \( r \) the first-order approximation might not be very satisfactory. Yet, one could additionally invoke the mean value theorem and consider average derivatives, i.e., the average markup and consequently average revenue shares of variable factors between firm \( r \) and firm \( i \), to improve the quality of the approximation. An alternative approach, that we fully develop below, consists instead in employing time changes. In this respect our findings based on the benchmark generalized CES case show a strong time persistence of the model’s fundamentals: productivity shocks, demand shocks and markups. The small time changes should provide a reasonably good base for our first-order approximation. However, the drawback of this approach is that we can only measure time changes of demand shocks within a firm. As for productivity and markups we can instead measure their level for all firms and years.

In what follows we use, for example, \( \Delta r_i = r_{it} - r_{i,t-1} \) for revenue. Equations (6) and (7) still hold in this broader setting and so, by proceeding as before, we get from (20), which is the equivalent of (8), to the following expression:

\[ \frac{\text{LHS}_{it}}{s_{Mat}} = \frac{\Delta r_{it} - s_{Lit} (\Delta k_{it} - \Delta k_{it}) - s_{Mat} (\Delta m_{it} - \Delta k_{it})}{s_{Mat}} = \frac{\gamma}{\alpha_M} \Delta k_{it} + \frac{1}{\alpha_M} (\Delta m_{it} + \Delta \lambda_{it}) \]  

(21)
which is the equivalent of (9). Combining (7) and (20) we then have:

\[
\Delta \lambda_{it-1} = \Delta r_{it-1} \frac{\alpha_M}{s_{Mit-1}} - \Delta q_{it-1}
\]

while plugging (7) and (20) into (21) and re-arranging yields:

\[
\Delta a_{it-1} = \alpha_M \widehat{LHS}_{it-1} - \gamma \Delta k_{it-1} - \left( \Delta r_{it-1} \frac{\alpha_M}{s_{Mit-1}} - \Delta q_{it-1} \right).
\]  (22)

Finally, by substituting the last two expressions as well as (10) into (21) we obtain:

\[
\widehat{LHS}_{it} = \gamma \frac{\alpha}{\alpha_M} \Delta k_{it} + \phi_a \frac{\gamma}{\alpha_M} \Delta k_{it-1} + \left( \phi_a - \phi_a \right) \left( \Delta r_{it-1} \frac{\alpha_M}{s_{Mit-1}} - \Delta q_{it-1} \right) + \frac{1}{\alpha_M} \left( \Delta \nu_{ait+1} + \nu_{\Delta \lambda_{it}} \right).
\]  (23)

which is the equivalent of (13). As in the case of (13), equation (23) can be estimated using a linear regression setting and an appropriate change in variables. However, contrary to (13), we cannot simply use OLS here because, for example, \( \nu_{ait-1} \) (which is included in \( \Delta \nu_{ait} \)) is correlated with \( \widehat{LHS}_{it-1} \), \( \Delta r_{it-1} \) and \( \Delta q_{it-1} \). A simple way to solve this issue is to instrument \( \widehat{LHS}_{it-1} \), \( \Delta r_{it-1} \) and \( \Delta q_{it-1} \) with their respective time lags.

Estimation of (23) delivers \( \hat{\beta} \) and \( \hat{\phi}_a \) that can be used to get an estimate of \( \gamma \) along the lines of what we show in Section 2. To do so, however, we first need to time-difference the Cobb-Douglas production constraint:

\[
\Delta q_{it} = \alpha_L (\Delta l_{it} - \Delta k_{it}) + \alpha_M (\Delta m_{it} - \Delta k_{it}) + \gamma \Delta k_{it} + \Delta a_{it}
\]

and make use of (6), (7), (10) and (22) to get to:

\[
\Delta q_{it} = \frac{\gamma}{\beta} \frac{s_{Lit}}{s_{Mit}} (\Delta l_{it} - \Delta k_{it}) + \frac{\gamma}{\beta} (\Delta m_{it} - \Delta k_{it}) + \gamma \Delta k_{it} + \Delta a_{it}
\]

\[
+ \phi_a \frac{\gamma}{\beta} \widehat{LHS}_{it-1} - \phi_a \gamma \Delta k_{it-1} - \phi_a \left( \Delta r_{it-1} \frac{\gamma}{\beta s_{Mit-1}} - \Delta q_{it-1} \right) + \Delta \nu_{ait}.
\]  (24)

(24) is the equivalent of (15) and can be manipulated in a similar fashion to estimate \( \gamma \) via a linear regression where the only covariate is instrumented with, for example, \( \Delta k_{it} \) based on the moment condition \( E \{ \Delta \nu_{ait} \Delta k_{it} \} = 0 \). Productivity shocks and markups can in turn be computed as before:

\[
\hat{a}_{it} = q_{it} - \frac{\gamma}{\beta} \frac{s_{Lit}}{s_{Mit}} (l_{it} - k_{it}) - \frac{\gamma}{\beta} (m_{it} - k_{it}) - \hat{k}_{it}
\]
\[ \hat{\mu}_{it} = \frac{\hat{\gamma}}{\hat{\beta}_{SMit}}. \]

However, as far as demand shocks are concerned, only their change over time within a firm can be measured:

\[ \Delta \lambda_{it} = \frac{\hat{\gamma}}{\hat{\beta}_{SMit}} \Delta r_{it} - \Delta q_{it}. \]

### A.2 Exact procedure with non log-linear demand

Here we discuss how our model can be extended without resorting to any linear approximation by fully specifying an alternative preference structure and work out the corresponding algebra for the estimation equations. We are particularly interested in a flexible structure leading to a non-log linear demand. One reason why one might choose such a structure is that it allows for markups varying with equilibrium quantity. Specifically, we look here at an additively separable utility function shaped like the Gaussian CDF:\(^{18}\)

\[ U(\tilde{q}) = \int_{i \in I} \Phi(\tilde{q}_i, \beta_0, \beta_1, \beta_2) \, di \]

where \( \Phi(\cdot) \) is the Gaussian cdf, i.e.,

\[ \Phi(\tilde{q}_i) = u(\tilde{q}_i) = \int_{-\infty}^{\tilde{q}_i} \phi(\tilde{q}_i) \, d\tau. \]

The inverse demand function for firm \( i \) at time \( t \) is consequently:

\[ \frac{\phi(\tilde{q}_{it})}{Q_{it} \kappa_t} = P_{it} \]

where \( \phi(\tilde{q}_{it}) \) is the Gaussian PDF. In what follows we set \( \phi(\tilde{q}_{it}) = \exp(-\beta_2^2 \tilde{q}_{it}^3 + \beta_1 \tilde{q}_{it} + \beta_0) \) but could have equally used more or less involved formulations.

The first thing to note is that the Gaussian utility as specified above implies a downward sloping demand curve for \( \beta_1 < 1 \). Indeed:

\[ \frac{\partial p_{it}}{\partial q_{it}} = -3\beta_2^2 \tilde{q}_{it}^2 + \beta_1 - 1 < 0 \text{ for } \beta_1 < 1 \]

Moving forward we have:\(^{19}\)

\[ \frac{\partial \ln \phi}{\partial \tilde{q}_{it}} = -3\beta_2^2 \tilde{q}_{it}^2 + \beta_1 = \frac{\partial p_{it}}{\partial q_{it}} + 1 = \frac{1}{\mu_{it}}, \]

as well as:

---

\(^{18}\)See Berhold (1973) for further discussion of the Gaussian CDF as a utility function.

\(^{19}\)We ignore terms constant across firms in the following.
\[ r_{it} = \ln \phi (q_{it}). \]

Using the definition \( \chi_{it} = s_{Liit} (l_{it} - k_{it}) + s_{Miit} (m_{it} - k_{it}) \) we obtain:

\[ \frac{1}{3} \frac{\partial \ln \phi}{\partial q_{it}} \tilde{q}_{it} = -\beta_2 \tilde{q}_{it}^3 + \frac{1}{3} \beta_1 \tilde{q}_{it} = r_{it} - \frac{2}{3} \beta_1 \tilde{q}_{it} \]

\[
\Rightarrow r_{it} = \left( \frac{1}{3} \frac{\partial \ln \phi}{\partial q_{it}} + \frac{2}{3} \beta_1 \right) \tilde{q}_{it} = \left( \frac{1}{3} \frac{\mu_{it}}{\mu_{iit}} + \frac{2}{3} \beta_1 \right) \tilde{q}_{it}
\]

\[
= \left( \frac{1}{3} + \frac{2}{3} \beta_1 \mu_{it} \right) \chi_{it} + \left( \frac{1}{3} \frac{\mu_{it}}{\beta_1} + \frac{2}{3} \beta_1 \beta_2 k_{it} + \left( \frac{1}{3} \frac{\mu_{it}}{\beta_1} + \frac{2}{3} \beta_1 \right) \beta_1 (a_{it} + \lambda_{it}) \right) \] 

\[
\Rightarrow \frac{r_{it}}{\frac{1}{3} \frac{\mu_{it}}{\beta_1} + \frac{2}{3} \beta_1} = \frac{1}{3} + \frac{2}{3} \beta_1 \mu_{it} \chi_{it} + \beta_1 \gamma k_{it} + \beta_1 (a_{it} + \lambda_{it}).
\]

Further note that:

\[
\frac{1}{3} + \frac{2}{3} \beta_1 \mu_{it} = \frac{1}{3} + \frac{2}{3} \beta_1 \mu_{it} = \beta_1 \frac{\alpha_m}{s_{Miit}}.\] Hence we can write:

\[
\frac{r_{it}}{\frac{1}{3} \frac{\mu_{it}}{\beta_1} + \frac{2}{3} \beta_1} = \beta_1 \alpha_m \frac{\chi_{it}}{s_{Miit}} + \beta_1 \gamma k_{it} + \beta_1 (a_{it} + \lambda_{it}),
\]

which implies:

\[
r_{it} = \left( \frac{1}{3} \frac{\mu_{it}}{\beta_1} + \frac{2}{3} \beta_1 \right) (q_{it} + \lambda_{it}) \tag{25}
\]

Equation (25) is the equivalent if (8). Building on the same logic utilized for (11) and (12) one finally gets:

\[
\begin{align*}
\text{LHS}_{it} &= \beta_1 \gamma k_{it} + \phi_a \text{LHS}_{it-1} - \phi_a \gamma k_{it-1} \\
&\quad + \left( \phi_a - \phi_a \right) \beta_1 k_{it-1} - \beta_1 (a_{it} + \lambda_{it}).
\end{align*} \tag{26}
\]

where \( \text{LHS}_{it} = \frac{s_{Miit}}{s_{Miit+1}} \). Estimation of the various parameters in (26) can be carried by non-linear GMM, as in De Loecker and Warzynski (2012), by considering the error term \( u_{it} = \beta_1 (v_{ait} + \nu_{ait}) \) is a function of some data as well as of the parameters and build on the following moment conditions: \( E [k_{it} u_{it}] = E [k_{it-1} u_{it}] = E [m_{it-1} u_{it}] = E [l_{it-1} u_{it}] = E [q_{it-1} u_{it}] = E [r_{it-1} u_{it}] = 0 \). Parallel to the generalized CES case one can avoid exploiting parameters’ constraints and extract some reduced-form parameters including \( \beta_1 \alpha_m \) and \( \beta_1 \gamma \) as well as \( \phi_a \). In the very same way we recover \( \gamma \) in the generalized CES case via a second step estimation based on the quantity equation we can, by using estimates of \( \beta_1 \alpha_m \), \( \beta_1 \gamma \) and \( \phi_a \), write
the quantity equation as a linear expression involving only one unknown parameter ($\beta_1$) and one right-hand side variables. Therefore, we can use a simple IV strategy based on the moment condition $E[k_{it}v_{ait}] = 0$ to identify $\beta_1$ and so $\alpha_M$, $\gamma$ and ultimately productivity, demand and markup shocks.

Notice that

$$\frac{1}{\mu_{it}} = -3\beta_2^2 q_{it}^2 + \beta_1$$

Hence

$$\partial \left( -3\beta_2^2 q_{it}^2 + \beta_1 \right) \over \partial q_{it} = -6\beta_2^2 q_{it}$$

while in the simple log-linear form the markup does not depend on equilibrium quantity.

**B More general production functions**

Here we show how we can introduce more flexible production functions. In particular we look at a (homogenous) translog form, i.e., our production function takes the form:

$$q_{it} = a_{it} + \sum_{X \in \{M, L, K\}} \left[ \alpha_X \ln X_{it} + \frac{1}{2} \alpha_{XX} \ln (X_{it})^2 \right] + \alpha_{MK} \ln M_{it} \ln K_{it} + \alpha_{ML} \ln M_{it} \ln L_{it} + \alpha_{LK} \ln L_{it} \ln K_{it}.$$  

Note that

$$\frac{\partial q_{it}}{\partial m_{it}} = \alpha_M + \alpha_{MM} m_{it} + \alpha_{MK} k_{it} + \alpha_{ML} l_{it}$$

$$\frac{\partial q_{it}}{\partial l_{it}} = \alpha_L + \alpha_{LL} l_{it} + \alpha_{LK} k_{it} + \alpha_{ML} m_{it}$$

$$\gamma - \frac{\partial q_{it}}{\partial m_{it}} - \frac{\partial q_{it}}{\partial l_{it}} = \alpha_K + \alpha_{KK} k_{it} + \alpha_{MK} m_{it} + \alpha_{LK} l_{it}$$

where the last equation follows from the homogeneity assumption (as before $\gamma$ represents the returns to scale).

Note that

$$\frac{\partial q_{it}}{\partial m_{it}} m_{it} + \frac{\partial q_{it}}{\partial l_{it}} l_{it} + \left( \gamma - \frac{\partial q_{it}}{\partial m_{it}} - \frac{\partial q_{it}}{\partial l_{it}} \right) k_{it}$$

$$= \alpha_M m_{it} + \alpha_L l_{it} + \alpha_K k_{it} + \alpha_{MM} m_{it}^2 + \alpha_{LL} l_{it}^2 + \alpha_{KK} k_{it}^2 + 2\alpha_{MK} k_{it} m_{it} + 2\alpha_{ML} m_{it} l_{it} + 2\alpha_{LK} l_{it} k_{it}$$

$$= q_{it} - a_{it} + \frac{1}{2} \alpha_{MM} m_{it}^2 + \frac{1}{2} \alpha_{LL} l_{it}^2 + \frac{1}{2} \alpha_{KK} k_{it}^2 + \alpha_{MK} k_{it} m_{it} + \alpha_{ML} m_{it} l_{it} + \alpha_{LK} l_{it} k_{it}$$
By substituting (10) to (12) into (27) (while replacing the old 
manipulating:

In this setting (8) holds and so by adding 

so that

From the first order conditions

\[ s_{Mlt} \mu_{it} = \frac{\partial q_{it}}{\partial m_{it}}, \quad s_{Llt} \mu_{it} = \frac{\partial q_{it}}{\partial l_{it}} \]

so that

\[ q_{it} = s_{Mlt} \mu_{it} m_{it} + s_{Llt} \mu_{it} l_{it} + (\gamma - s_{Mlt} \mu_{it} - s_{Llt} \mu_{it}) k_{it} \]

\[ -\frac{1}{2} \alpha_{MM} m_{it}^2 - \frac{1}{2} \alpha_{LL} l_{it}^2 - \frac{1}{2} \alpha_{KK} k_{it}^2 - \alpha_{MK} k_{it} m_{it} - \alpha_{ML} m_{it} l_{it} - \alpha_{LK} l_{it} k_{it} + a_{it}. \]

In this setting (8) holds and so by adding \( \lambda_{it} \) on both sides and dividing by \( 1/\mu_{it} \) one gets after manipulating:

\[ \text{LHS}_{it} \frac{\partial q_{it}}{\partial m_{it}} = (q_{it} + \lambda_{it}) - \mu_{it} [s_{Llt} (l_{it} - k_{it}) + s_{Mlt} (m_{it} - k_{it})] \]

\[ = \gamma k_{it} + \alpha_{MM} m_{it}^2 + \alpha_{LL} l_{it}^2 + \alpha_{KK} k_{it}^2 + \alpha_{MK} k_{it} m_{it} + \alpha_{ML} m_{it} l_{it} + \alpha_{LK} l_{it} k_{it} + (a_{it} + \lambda_{it}), \]

where \( \text{LHS}_{it} \) is the same as in (9), i.e., a function of observables and \( \partial q_{it}/\partial m_{it} = \alpha_M + \alpha_{MM} m_{it} + \alpha_{MK} k_{it} + \alpha_{ML} l_{it} \), i.e., a function of some parameters as well as \( m_{it}, l_{it} \) and \( k_{it} \).

By substituting (10) to (12) into (27) (while replacing the old \( \alpha_M \) with \( \partial q_{it}/\partial m_{it} \) and dividing both sides by \( \gamma \) we get:

\[ \text{LHS}_{it} \frac{\partial q_{it}}{\partial m_{it}} \frac{1}{\gamma} = e_{it} - \phi_a e_{it-1} + \partial q_{it}/\partial m_{it} \frac{1}{\gamma} (\phi_a \text{LHS}_{it-1}) \]

\[ + \partial q_{it}/\partial m_{it} \frac{1}{\gamma} (\phi_a - \phi_a) \left( \frac{r_{it-1}}{s_{Miit-1}} - \frac{1}{\partial q_{it}/\partial m_{it}} q_{it-1} \right) + u_{it} \]

where \( e_{it} = k_{it} + \frac{\alpha_{MM}}{\gamma} m_{it}^2 + \frac{\alpha_{LL}}{\gamma} l_{it}^2 + \frac{\alpha_{KK}}{\gamma} k_{it}^2 + \frac{\alpha_{MK}}{\gamma} k_{it} m_{it} + \frac{\alpha_{ML}}{\gamma} m_{it} l_{it} + \frac{\alpha_{LK}}{\gamma} l_{it} k_{it} \) and \( u_{it} = \frac{1}{\gamma} (v_{ait} + v_{ait}). \)

Estimation of the various parameters in (28) can be carried by non-linear GMM, as in De Loecker and Warzynski (2012), by considering that \( tu_{it} \) is a function of some data as well as of the parameters and build on moment conditions like \( E [k_{it} u_{it}] = E [k_{it-1} u_{it}] = E [m_{it-1} u_{it}] = E [l_{it-1} u_{it}] = 0 \) as well as \( E [m_{it-1} u_{it}] = E [m_{it-1} k_{it-1} u_{it}] = E [m_{it-1} l_{it-1} u_{it}] = E [k_{it-1} l_{it-1} u_{it}] = 0 \) and so on and do forth. Considering moments up to \( t - 2 \) (\( t - 1 \)) there are 30 (13) such moments conditions that can be
exploited. As in the Coob-Douglas case it perhaps best not to exploit parameters’ constraints (this mean for example estimating $\frac{\partial q_{it}}{\partial m_{it}}$ rather than trying to separately identify $\alpha_{MM}$ and $\gamma$ from the revenue equation) and extract some reduced form parameters to be used in a second stage regression based on the quantity equation.

For the quantity equation we have:

$$q_{it} = \frac{\partial q_{it}}{\partial m_{it}} s_{M_{it}} + \gamma k_{it} - \frac{1}{2} \alpha_{MM} m_{it}^2 - \frac{1}{2} \alpha_{LL} l_{it}^2 - \frac{1}{2} \alpha_{KK} k_{it}^2 - \alpha_{MK} m_{it} k_{it} - \alpha_{ML} m_{it} l_{it} - \alpha_{LK} l_{it} k_{it} + a_{it},$$  \hspace{1cm} (29)

where $\chi_{it} = s_{L_{it}} (l_{it} - k_{it}) + s_{M_{it}} (m_{it} - k_{it})$. All of the parameters in (30) have been identified up to the scaling $\gamma$ in the previous stage and we can write it as:

$$q_{it} = \gamma z_{it} + a_{it},$$ \hspace{1cm} (30)

where:

$$z_{it} = \frac{\partial q_{it}}{\partial m_{it}} \frac{1}{\gamma} \chi_{it} + k_{it} - \frac{1}{2} \alpha_{MM} m_{it}^2 - \frac{1}{2} \alpha_{LL} l_{it}^2 - \frac{1}{2} \alpha_{KK} k_{it}^2 - \frac{\alpha_{MK}}{\gamma} m_{it} k_{it} - \frac{\alpha_{ML}}{\gamma} m_{it} l_{it} - \frac{\alpha_{LK}}{\gamma} l_{it} k_{it} + a_{it}.$$

As in the Cobb-Douglas case we can further substitute for $a_{it}$ using (12) (while replacing the old $\alpha_M$ with $\frac{\partial q_{it}}{\partial m_{it}}$) and use the moment condition $E[k_{it} \nu_{ait}] = 0$ on a simple linear model with a single regressor to identify $\gamma$ and ultimately productivity, demand and markup shocks.

**C More general processes for $a$ and $\lambda$**

Our model can be easily extended to non-linear Markov processes for $a$ and $\lambda$ as well as to the presence of time-invariant unobservable heterogeneity. Consider, for example, the former case and in particular:

$$a_{it} = \phi_{1a} a_{it-1} + \phi_{2a} a_{it-1}^2 + \nu_{ait}$$

$$\lambda_{it} = \phi_{1\lambda} \lambda_{it-1} + \phi_{2\lambda} \lambda_{it-1}^2 + \nu_{\lambda it}.$$ \hspace{1cm} (31)

By substituting (11), (12) and (31) into (9) we obtain:
\[ LHS_{it} = \frac{\gamma}{\alpha_M} k_{it} + \phi_{1a} LHS_{it-1} - \phi_{1a} \frac{\gamma}{\alpha_M} k_{it-1} \\
+ (\phi_{1\lambda} - \phi_{1a}) \left( \frac{r_{it-1}}{SM_{it-1}} - \frac{1}{\alpha_M} q_{it-1} \right) \\
+ (\phi_{2\lambda} - \phi_{2a}) \left( \alpha_M \left( \frac{r_{it-1}}{SM_{it-1}} \right)^2 + \frac{1}{\alpha_M} (q_{it-1})^2 - 2 \left( \frac{r_{it-1}}{SM_{it-1}} q_{it-1} \right) \right) \\
+ \phi_{2a} \left( \alpha_M (LHS_{it-1})^2 + \frac{\gamma^2}{\alpha_M} (k_{it-1})^2 - 2 \gamma LHS_{it-1} k_{it-1} - 2 \alpha_M (LHS_{it-1}) \frac{r_{it-1}}{SM_{it-1}} \right) \\
+ \phi_{2a} \left( 2 \gamma (k_{it-1}) \frac{r_{it-1}}{SM_{it-1}} + 2 (LHS_{it-1} q_{it-1}) - 2 \frac{\gamma}{\alpha_M} (k_{it-1} q_{it-1}) \right) \right) + \frac{1}{\alpha_M} (v_{ait} + v_{\lambda it}) . \] 

(32) can be used to estimate \( \frac{\gamma}{\alpha_M} \equiv \beta , \phi_{1a} \) and \( \phi_{2a} \) by a suitable linear regression with a change in variables and some reduced-form parameters. In turn these estimates could be employed in the corresponding expression of (15):

\[ q_{it} = \frac{\gamma}{\beta} \frac{s_{it}}{SM_{it}} (l_{it} - k_{it}) + \frac{\gamma}{\beta} (m_{it} - k_{it}) + \gamma k_{it} \\
+ \hat{\phi}_{1a} \frac{\gamma}{\beta} LHS_{it-1} - \hat{\phi}_{1a} \gamma k_{it-1} - \hat{\phi}_{1a} \left( \frac{r_{it-1}}{\beta SM_{it-1}} - q_{it-1} \right) \tag{33} \\
+ \hat{\phi}_{2a} \left( \frac{\gamma^2}{\beta} (LHS_{it-1})^2 + \hat{\phi}_{2a} \gamma^2 (k_{it-1})^2 + \hat{\phi}_{2a} \frac{\gamma^2}{\beta} (r_{it-1})^2 \right) + \hat{\phi}_{2a} (q_{it-1})^2 \\
- 2 \hat{\phi}_{2a} \frac{\gamma^2}{\beta} LHS_{it-1} k_{it-1} - 2 \hat{\phi}_{2a} \left( \frac{\gamma^2}{\beta} LHS_{it-1} \frac{r_{it-1}}{SM_{it-1}} + 2 \hat{\phi}_{2a} \frac{\gamma}{\beta} LHS_{it-1} q_{it-1} \right) \\
+ 2 \hat{\phi}_{2a} \frac{\gamma^2}{\beta} k_{it-1} \frac{r_{it-1}}{SM_{it-1}} - 2 \hat{\phi}_{2a} \gamma k_{it-1} q_{it-1} + v_{ait} . \]

from which the \( \gamma \) parameter can obtained by a suitable linear regression where the dependent variable is \( q_{it} - \hat{\phi}_{1a} q_{it-1} - \hat{\phi}_{2a} (q_{it-1})^2 \) and the right-hand side variables are grouped into two sets: one in which the only unknown coefficient is \( \gamma \) and the other where the only unknown coefficient is \( \gamma^2 \). As in the baseline case instrumenting is needed.

The case of time-invariant unobservable heterogeneity is easier to handle. In this scenario we have:

\[ a_{it} = \hat{\phi}_a a_{it-1} + u_{ait} + v_{ait} \]
\[ \lambda_{it} = \phi_\lambda \lambda_{it-1} + u_{\lambda it} + v_{\lambda it} . \] 

(34)

By substituting (11), (12) and (34) into (9) we obtain:

\[ \text{As in the baseline case } \beta \text{ and } \phi_{it} \text{ can be directly obtained from, respectively, the coefficients of } k_{it} \text{ and } LHS_{it-1} \text{ with no need to exploit reduced-form parameters constraints. As for } \phi_{2a} \text{, this can be obtained as } 1/2 \text{ times the coefficient of the interaction between } LHS_{it-1} \text{ and } q_{it-1}. \]
\[ LHS_{it} = \frac{\gamma}{\alpha_M} k_{it} + \phi_a LHS_{it-1} - \phi_a \frac{\gamma}{\alpha_M} k_{it-1} \]
\[ + \left( \phi_\lambda - \phi_a \right) \left( \frac{r_{it-1}}{s_{it-1}} - \frac{1}{\alpha_M} q_{it-1} \right) + \frac{1}{\alpha_M} (u_{ai} + u_{\lambda_{it}}) + \frac{1}{\alpha_M} (v_{ai} + v_{\lambda_{it}}). \]  

(35)

(35) can be transformed into a linear regression model similar to (14) with the only difference being that, the presence of an unobservable time-invariant component \(\frac{1}{\alpha_M} (u_{ai} + u_{\lambda_{it}})\) correlated with regressors in (35), calls for the use of, for example, a within estimator rather than OLS. The same argument applies to quantity equation (16).

D Aggregation

We now provide an example in which it makes sense to aggregate quantities produced of different products within a firm while using average log prices (across firms within a product) as weights. In terms of our data a product has to be thought as an 8-digit Prodcom code produced by a firm belonging to a given industry, i.e., at the 3-digit unit of measurement level.

Suppose that firm \(i\) produces many products indexed by \(j\) and that the log production function of product \(j\) by firm \(i\) can be simplified as \(q_{ij} = q_i + s^q_j\) where \(s^q_j\) is an Hicks-neutral shifter specific to the product and constant across firms. Further assume that the log demand shock corresponding to product \(j\) produced by firm \(i\) is \(\lambda_{ij} = \lambda_i + s^\lambda_j\) where \(s^\lambda_j\) is specific to the product and constant across firms. Now impose markups \(\mu_{ij} = \mu_i\). We thus get:

\[ r_{ij} = p_{ij} + q_{ij} = \frac{1}{\mu_i} (q_{ij} + \lambda_{ij}) = \frac{1}{\mu_i} (s_j + q_i + \lambda_i) = \frac{1}{\mu_i} (\bar{q}_{ij} + \lambda_i), \]

where \(s_j = s^q_j + s^\lambda_j\) and \(\bar{q}_{ij} = s_j + q_i\). This equation shows that, within our assumptions, everything works as if the firm was producing identical products, i.e., having the same productivity, demand and markups shocks as well as technology constraint, in different quantities \(\bar{q}_{ij}\). The problem is that \(q_{ij}\) is directly observable in our data while \(\bar{q}_{ij}\) is not. Yet, from the above equation we get:

\[ p_{ij} = r_{ij} - q_{ij} = \frac{1}{\mu_i} (\bar{q}_{ij} + \lambda_i) - q_i - s^q_j = \frac{1}{\mu_i} (q_i + \lambda_i) - q_i + \frac{1}{\mu_i} s_j - s^q_j. \]

We finally posit \(E[p_{ij} | i \in I^j] = a + bs_j - s^q_j\) where \(I^j\) is the set of firms \(i\) producing product \(j\). This would be automatically satisfied if all firms were producing all products. On a broader basis, this amounts to assume that the distributions of productivity, markups and demand shocks corresponding to firms selling a given product are similar across the 8-digit products belonging to a given industry.

We thus allow for the distributions of firm-level productivity, markups and demand shocks to be different across industries.

The above assumption implies that, by multiplying the average (across firms within a product) log price \(p_{ij}\) observed in the data by the observed \(q_{ij}\) we get a monotonous transformation of \(\bar{q}_{ij}\):
\[ \mathbb{E} \left[ p_{ij} | i \in I^j \right] q_{ij} = a + bs_j + q_i. \]

that we can use to quantify parameters. Note that using a monotonous transformation of \( \tilde{q}_{ij} \) rather than \( \tilde{q}_{ij} \) is irrelevant for the purpose of our model given our extensive use of linear regressions.