DISCUSSION PAPER SERIES

No. 11058

UNRAVELING FIRMS: DEMAND, PRODUCTIVITY AND MARKUPS HETEROGENEITY

Emanuele Forlani, Ralf Martin, Giordano Mion and Mirabelle Muûls

INDUSTRIAL ORGANIZATION and INTERNATIONAL TRADE AND REGIONAL ECONOMICS

ISSN 0265-8003

UNRAVELING FIRMS: DEMAND, PRODUCTIVITY AND MARKUPS HETEROGENEITY

Emanuele Forlani, Ralf Martin, Giordano Mion and Mirabelle Muûls

Discussion Paper No. 11058
January 2016
This revision December 2017

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: (44 20) 7183 8801
www.cepr.org

UNRAVELING FIRMS: DEMAND, PRODUCTIVITY AND MARKUPS HETEROGENEITY†

Abstract

We develop a novel econometric framework that simultaneously allows recovering heterogeneity in demand, quantity TFP and markups across firms while leaving the correlation among the three unrestricted. We accomplish this by explicitly introducing demand heterogeneity and by systematically exploiting assumptions used in previous productivity estimation approaches. In doing so, we provide an exact decomposition of revenue productivity in terms of the underlying heterogeneities thus bridging the gap between quantity and revenue productivity estimations. We use Belgian firms production data to quantify TFP, demand and markups and show how they are correlated with each other, across time and with measures obtained from other approaches. We also show how and to what extent our three dimensions of heterogeneity provide deeper and sharper insights on two questions: firm response to increasing import competition from China and the productivity advantage of importers.

JEL Classification: D24, F14, L11 and L25

Keywords: Demand, Productivity, Markups, Production function estimation, Import competition, China, Import status

Emanuele Forlani emanuele.forlani@unibo.it

University of Bologna and Centro Studi Luca d'Agliano

Ralf Martin r.martin@imperial.ac.uk

Imperial College Business School and Grantham Institute, Imperial College London, and CEP

Giordano Mion g.mion@sussex.ac.uk

University of Sussex, CEP, CESifo and CEPR

Mirabelle Muûls m.muuls@imperial.ac.uk

Imperial College Business School and Grantham Institute, University College London, and CEP

† We thank Andrea Ariu, Andrew Bernard, Ingo Borchert, Rosario Crino, Jan De Loecker, Catherine Fuss, Marc Melitz, Luca David Opromolla, Chad Syverson, Ilke Van Beveren and John Van Reenen as well as participants at various seminars and workshops and an anonymous referee from the National Bank of Belgium. 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.

1 Introduction

Economists are interested in estimating firm-level productivity in a range of fields. These estimates are often used as inputs in a number of applications such as the firm size distribution, firm survival and growth, self-selection of firms into trade activity and the extensive and intensive margins of trade to name a few. In the literature, the most commonly used approach to estimate productivity 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 at least 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. Such revenue-based measures of TFP look quite different from quantity-based ones (Foster et al., 2008). Second, a well known issue is the endogeneity of production factors used as explanatory variables (Olley and Pakes, 1996). Third and more importantly, firms could be heterogeneous in dimensions other than technical efficiency. In this respect the IO literature on demand systems (Ackerberg et al., 2007) points to substantial heterogeneity in both markups and consumers' willingness to pay for the products sold by different firms. For example, the presence of vertical and horizontal product differentiation means that firms selling otherwise similar products face rather different demands. At the same time market power variations, due to product quality or technical efficiency, could substantially affect the markup that firms can charge. Moreover, markups could also vary because of factors unrelated to either product attributes or efficiency; e.g. some firms could have a better market access than others because of spatial differentiation and trade costs. Being able to account for these different dimensions and their interconnections is important for several reasons.

First of all it is crucial in order to correctly measure TFP. In this respect higher measured TFP is typically seen as welfare improving. However, conventional measures of TFP conflate actual TFP with demand and markups heterogeneity which may lead to different welfare implications. Second, being able to actually quantify dimensions other than TFP matters from both a welfare and a policy point of view. From a welfare perspective it is, for example, of great value to assess the impact on firm markups of a trade integration episode or market size expansion. Some recent theoretical papers have indeed revisited the relationship between market size, markups and welfare and questioned the pervasiveness of the so called "pro-competitive effects" (Dhingra and Morrow, 2012 and Zhelobodko et al., 2012). Furthermore, being able to disentangle demand heterogeneity from efficiency is important for policy matters and in particular to understand where the competitiveness of a firm or an industry comes from and then target interventions accordingly.

This paper's contribution is to address these issues in a comprehensive way. We explicitly introduce demand heterogeneity across firms and model it as differences in the appeal/quality of a firm's products as in Hottman et al. (2016). In what follows we thus refer to demand heterogeneity and product appeal interchangeably. We then develop a model and estimation strategy that allow recovering productivity, demand, and markups heterogeneity across firms, while leaving the correlation among the three unrestricted. In doing so, we also provide an exact decomposition of revenue productivity in terms of the underlying heterogeneities so bridging the gap between quantity and revenue productivity estimations. We deal with the endogeneity issues related to the estimation of the production function in the standard way, i.e, by imposing a Markov process for productivity while assuming capital in predetermined. We then build upon Hall (1986)'s result relating markups to output elasticity and variable factors revenue shares to recover markups. We finally impose a restriction on the elasticity of revenue with respect to product appeal that amounts assuming consumers buy quantities that enter utility as quantity times product appeal. Our framework is rich enough to allow for multi-product firms, alternative hypotheses on preferences and market structure as well as on the production function and processes for productivity and demand. At the same time, our framework is parsimonious enough to allow retrieving productivity, demand, and markups heterogeneity with relatively little information compared to demand systems models.[1] These features provide a wide scope of applications to our framework.

We apply our econometric framework to Belgian manufacturing firms and use information on both the quantity and the value of production over the period 1996-2007 to quantify our model. We first document that demand factors display at least as much variability across firms as quantity TFP. We further show that productivity and demand heterogeneity are very strongly and negatively correlated in all industries. This finding is robust to using an alternative estimation approach based upon De Loecker et al. (2016) and it is suggestive of a trade-off between the quality of a firm's products and their production cost as suggested in Ackerberg et al. (2007). 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

[1] Demand system models have very rich structures and allow for consumer and product specific elasticities of demand. However, they require detailed information on product and consumer characteristics as well as suitable instruments (like cost shifters) for identification. See, for example, Berry et al. (2004) and Roberts et al. (2016) The high data requirements of these models are such that their application is usually limited to specific industries and contexts. By contrast, our simpler and more parsimonious framework only requires information on product prices, quantities and inputs and does not need any additional instrument.

Mercedes plant. Rather, Mercedes and Nissan face quite different demands which leads to different prices as well as different markups. Both plants are profitable and perhaps generate a very similar revenue-based productivity. Yet, their business model is rather different and our framework allows pinning this down.

Another pattern worth noting is that differences in markups across firms are reasonably well explained (in terms of R^2) by differences in demand and productivity. More specifically, we find that more productive firms and/or firms selling more appealing products charge higher markups. However, there remains a fair amount of unexplained variation. In this respect, our framework allows for markups that are either entirely determined by productivity and demand heterogeneity, as well as predetermined inputs, or are also determined by other factors. We further show how, when correctly measured, revenue TFP exactly decomposes into the underlying dimensions of heterogeneity so bridging the gap between quantity TFP estimations and revenue TFP estimations. We also provide evidence that standard measures of revenue TFP used in the literature correlate well and in a meaningful way with our heterogeneities.

We finally assess how and to what extent these heterogeneities allow gaining deeper and sharper insights into two productivity-related questions: firm response to increasing import competition from China and the productivity advantage of importers. Considering the former, we show how changes in firm revenue productivity spurred by import competition from China materialize as the outcome of complex changes in quantity TFP, product appeal, markups and production scale. This in turns allows to better understand firm behaviour and margins of adjustment under competitive pressure and learn useful lessons that can applied to other contexts. By decomposing the revenue productivity advantage of importers, we further provide evidence that demand heterogeneity across firms is at least as important as quantity TFP in drawing the line between importing and non-importing firms. This suggests that international trade model à la Melitz (2003) or Antras and Helpman (2004) should pay more attention to demand heterogeneity while at the same time trade policies should be paying equal attention to production efficiency and quality/marketing issues.

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), Wooldridge (2009), Ackerberg et al. (2015) and De Loecker et al. (2016). In our estimation procedure we depart from the proxy variable approach and in particular from the problematic invertibility assumption. More specifically, we use both the revenue and quantity equations to recover technology parameters. Indeed, we are sufficiently explicit about demand to be able to explicitly write down revenue as a function of observables and the heterogeneities we allow for and use both the revenue and quantity equations to estimate technology parameters. However, we also show that using an alternative estimation strategy based on De Loecker et al. (2016) does not affect our results.

Our interest in demand heterogeneity is common to De Loecker (2011), Foster et al. (2008) and Jaumandreu and Yin (2017). De Loecker (2011) introduces demand heterogeneity in a revenue-based production function model while relying on standard CES preferences and a common markup across varieties. This allows substituting for prices and getting a tractable expression for firm revenue as a function of inputs, TFP and demand heterogeneity. Compared to our framework, De Loecker (2011) does not allow for different markups across varieties while needing some adequate proxies for demand shocks. By contrast, Foster et al. (2008) use data on both the quantity and the value of a firm's production in order to disentangle quantity TFP from demand heterogeneity. More specifically, they focus on homogeneous goods and recover production function coefficients from industry average cost shares. They subsequently estimate a demand system featuring demand heterogeneity measured as regression residuals and instrument firm price with firm TFP. Therefore, the key identifying assumption allowing them to disentangle productivity from demand is, besides imposing constant markups, that they are uncorrelated. In our framework we do not impose such assumptions and find productivity and demand heterogeneity to be very strongly correlated with each other. Jaumandreu and Yin (2017) provide a framework allowing for the presence of correlated demand and TFP heterogeneity. However, they do not observe quantities and so lay down a number of assumptions under which the two heterogeneities can be recovered from revenue and inputs data as well as data on demand shifters. Interestingly, they also find TFP and demand heterogeneity to be negatively correlated. Our approach requires less assumptions and provides, among others, what we believe is a more direct and compelling evidence about the negative correlation between TFP and demand heterogeneity.

The rest of the paper is organized as follows. Section 2 provides our econometric model while in Section 3 we describe our estimation strategy. We briefly discuss various extensions to the model in Section 4 that we fully develop in the Appendix. We present our data in Section 5 while Section 6 contains estimation results as well as descriptive statistics and correlations. In Section 7 we show how our framework can be used to get fresh insights into two productivity-related questions: firm response to increasing import competition from China and the productivity advantage of importers. Section 8 concludes.

2 The MULAMA model

We label our model MULAMA because of the names we give to the 3 heterogeneities we allow for: markups **MU**, demand **LAM**bda and productivity **A**. We present the model in Sections 2.1 to 2.3 while discussing identification in Section 2.4 and providing in Section 2.5 an exact decomposition of revenue productivity in terms of our heterogeneities.

2.1 Production

We index firms by i and time by t. In what follows we consider, for ease of exposition, a Cobb-Douglas production technology with 3 production factors: labour (L), materials (M) and capital (K). We consider the more involved Translog case in Appendix B. In line with the existing literature we assume capital to be a dynamic input that is predetermined in the short-run, i.e., current capital has been chosen in the past and cannot immediately adjust to current period shocks.[2] We further assume, as standard in the literature, that materials are a variable input free of adjustment costs. Concerning labor we could assume it is a variable input free of adjustment costs (a case we consider in Appendix E), or we could assume it is, very much like capital, predetermined in the short-run as in De Loecker et al. (2016), or we could also assume, following Ackerberg et al. (2015), it is a semi-flexible input.[3] In light of the features of the Belgian labor market we opt for the predetermined case.

We further assume firms are single-product, while relaxing this assumption in Appendix D, and minimize costs while taking the price of 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:[4]

$$\min_{M_{it}} \{M_{it}W_M\} \text{ s.t. } Q_{it} = A_{it}L_{it}^{\alpha_L}M_{it}^{\alpha_M}K_{it}^{\gamma-\alpha_L-\alpha_M},$$

where A_{it} is quantity TFP, γ characterizes returns to scale and A_{it} is observable to the firm (and influences her choices) but not to the econometrician. In what follows we refer to the Cobb-Douglas production technology as the quantity equation and denote with lower case the log of a variable (for example a_{it} denotes the natural logarithm of A_{it}). The quantity

[2] As described in Ackerberg et al. (2015) capital is often assumed to be a dynamic input subject to an investment process with the period t capital stock of the firm actually determined at period t-1. Intuitively, the restriction behind this assumption is that it takes a full period for new capital to be ordered, delivered, and installed.

[3] More precisely, in the semi-flexible case L_{it} is chosen by firm i at time $t-b$ ($0 < b < 1$), after K_{it} being chosen at $t-1$ but prior to M_{it} being chosen at t. In this case, one should expect L_{it} to be correlated with productivity shocks in t. Yet labour would not adjust fully to such shocks as materials do. The choice between predetermined and semi-flexible for L_{it} does not change the structure of the model and estimation procedure we provide below but only affects the set of moments used in the estimation. We highlight any differences later on.

[4] To simplify notation we ignore components that are constant across firms in a given time period as they will be controlled for by time dummies.

equation can thus be written as:

$$q_{it} = \alpha_L l_{it} + \alpha_M m_{it} + (\gamma - \alpha_L - \alpha_M) k_{it} + a_{it}. \tag{1}$$

First order conditions to the firm's cost minimization problem imply that:

$$W_M = \chi_{it} \frac{Q_{it}}{M_{it}} \alpha_M \tag{2}$$

where χ_{it} is a Lagrange multiplier.[5]

We can thus write the short-run cost function as:

$$C_{it} = M_{it} W_M = \chi_{it} Q_{it} \alpha_M = W_M \left(\frac{Q_{it}}{A_{it}} \right)^{\frac{1}{\alpha_M}} L_{it}^{-\frac{\alpha_L}{\alpha_M}} K_{it}^{-\frac{\gamma-\alpha_L-\alpha_M}{\alpha_M}}. \tag{3}$$

Marginal cost thus satisfies the following property:

$$\frac{\partial C_{it}}{\partial Q_{it}} = \frac{1}{\alpha_M} \frac{C_{it}}{Q_{it}}. \tag{4}$$

2.2 Markups and the productivity process

The above assumptions are enough to provide a simple rule to pin-down firm-level markups irrespective of the underlying demand and market structure faced by firms. This result, highlighted in Hall (1986) and implemented in De Loecker and Warzynski (2012) and De Loecker et al. (2016), simply requires cost-minimization of a variable input free of adjustment costs (materials for us) and price-taking behaviour on the inputs side (W_M is given to the firm). The proof in the general case goes as follows:

$$\frac{\partial C_{it}}{\partial Q_{it}} = \frac{\partial C_{it}}{\partial M_{it}} \frac{\partial M_{it}}{\partial Q_{it}} = W_{Mt} \frac{\partial M_{it}}{\partial Q_{it}}.$$

Now define the markup as:

$$\mu_{it} \equiv \frac{P_{it}}{\frac{\partial C_{it}}{\partial Q_{it}}}.$$

We thus have:

$$\frac{P_{it}}{\mu_{it}} = W_{Mt} \frac{\partial M_{it}}{\partial Q_{it}}.$$

Multiplying by Q_{it} and dividing by M_{it} on both sides:

$$\frac{P_{it} Q_{it}}{M_{it} \mu_{it}} = \frac{R_{it}}{M_{it} \mu_{it}} = W_{Mt} \frac{\partial M_{it}}{\partial Q_{it}} \frac{Q_{it}}{M_{it}} = W_{Mt} \frac{\partial m_{it}}{\partial q_{it}}.$$

[5] $\chi_{it} = \frac{W_M}{\alpha_M} Q_{it}^{\frac{1}{\alpha_M}-1} A_{it}^{-\frac{1}{\alpha_M}} L_{it}^{-\frac{\alpha_L}{\alpha_M}} K_{it}^{-\frac{\gamma-\alpha_L-\alpha_M}{\alpha_M}}$.

Re-arranging we finally have:

$$\mu_{it} = \frac{\frac{\partial q_{it}}{\partial m_{it}}}{\frac{W_{Mt}M_{it}}{R_{it}}} = \frac{\frac{\partial q_{it}}{\partial m_{it}}}{s_{Mit}}.$$

The simple rule to pin-down markups thus consists in taking the ratio of the output elasticity of materials ($\frac{\partial q_{it}}{\partial m_{it}}$) to the share of materials in revenue ($s_{Mit} \equiv \frac{W_{Mt}M_{it}}{R_{it}}$). In our baseline Cobb-Douglas case this becomes:

$$\mu_{it} = \frac{\alpha_M}{s_{Mit}}, \tag{5}$$

and can also be obtained by combining equations (2), (3) and (4). Such simple rule means that, if one imposes sufficient assumptions to be able to consistently estimate the parameters of the underlying production function (and so recover an estimate $\widehat{\frac{\partial q_{it}}{\partial m_{it}}}$), markups estimates can then be directly obtained without need to make explicit assumptions about firm demand and market structure. This is the strategy followed in De Loecker and Warzynski (2012) and De Loecker et al. (2016) by means of standard assumptions made within the proxy variable approach: existence and invertibility of a suitable conditional input demand for materials and a Markov process for productivity. We do not impose the former assumption, which will be discussed later on when comparing our framework to these frameworks, but do impose the latter. Again, for ease of exposition, we consider here the leading AR(1) case while describing how to handle richer cases in Appendix C.[6] More specifically we assume:

$$a_{it} = \phi_a a_{it-1} + \nu_{ait} \tag{6}$$

where ν_{ait} are iid and uncorrelated with past values of productivity.

As will be discussed in more details in the next Section, we do not need to impose existence and invertibility of a suitable conditional input demand for materials because we are sufficiently explicit about demand heterogeneity and market structure to write down revenue as a function of observables and the heterogeneities we allow for and use both such revenue equation and the quantity equation (1) to estimate technology parameters. At the same time this will allow us to further quantify demand heterogeneity.

2.3 Demand heterogeneity and market structure

Our key contribution is to allow for firms to face heterogeneous demands while being able to actually quantify this additional dimension of heterogeneity. In doing so, we do not impose a priori any restrictions on the correlations between demand, productivity and markups heterogeneity. This is in contrast to Foster et al. (2008) who instead impose a common markup

[6]In Appendix C we consider more general cases including non-linear terms, endogenous processes based on a simple extension of the model developed in Doraszelski and Jaumandreu (2013) as well as unobserved time-invariant heterogeneity; the latter being something the standard proxy variable approach cannot handle.

across firms and a zero correlation between demand and productivity.[7] In turn, this allows us to derive an exact decomposition of revenue productivity in terms of these heterogeneities, that we provide in Section 2.5, so bridging the gap between quantity TFP estimations and revenue TFP estimations.

In order to be able to quantify demand heterogeneity we need two elements. First, we have to spell out what we exactly mean by demand heterogeneity and what properties characterize it. This is what we do in this Section. Second, we need to impose some identification assumptions. This is what we do in Section 2.4. Moving to the first item, we assume that demand heterogeneity across firms is characterized by a measure of consumers' willingness to pay for a particular product (Λ_{it}) that is observable to the firm (and influences her choices) but not the econometrician. Later on we will also refer to Λ_{it} as product appeal or quality, for reasons that will become soon clear. To anticipate on this, products with higher Λ_{it} could be considered of a higher perceived quality by consumers because, if they were priced the same as products with lower Λ_{it}, consumers would buy more.

Let's first start from a property derived from profit maximization that will be useful in defining the scope of our model. Standard profit maximization (marginal revenue equal to marginal costs) implies that the elasticity of revenue with respect to quantity is one over the markup:

$$\frac{\partial r_{it}}{\partial q_{it}} = \underbrace{\frac{\partial R_{it}}{\partial Q_{it}}}_{\text{marginal revenue}} \frac{Q_{it}}{R_{it}} = \underbrace{\frac{\partial C_{it}}{\partial Q_{it}}}_{\text{marginal cost}} \frac{Q_{it}}{P_{it}Q_{it}} = \frac{\frac{\partial C_{it}}{\partial Q_{it}}}{P_{it}} = \frac{1}{\mu_{it}}, \tag{7}$$

where μ_{it} is the profit maximizing markup. This result comes from static profit maximization and holds under different assumptions about demand (representative consumer and discrete choice models) and market structure (monopolistic competition and various forms of oligopoly).

We impose that demand heterogeneity is such that the elasticity of revenue with respect to quantity is proportional to the elasticity of revenue with respect to Λ_{it}:

$$\frac{\partial r_{it}}{\partial \lambda_{it}} = \frac{b}{\mu_{it}} = b\,\frac{\partial r_{it}}{\partial q_{it}}. \tag{8}$$

where $b > 0$. We show here, and more in-depth in Appendix A, that this can be embedded into static profit maximization settings spanning over different demand and market structure assumptions. In the estimation procedure we impose, without loss of generality, $b = 1$ so that

[7] To be more specific, Foster et al. (2008) use their quantity TFP measure as an instrument for firm price in a regression where quantity is on the left hand-side, price is on the right-hand side and the regression residual is defined as demand heterogeneity. Clearly, this amounts to impose that the instrument (quantity TFP) is uncorrelated with the residual (demand heterogeneity).

(8) becomes:

$$\frac{\partial r_{it}}{\partial \lambda_{it}} = \frac{\partial r_{it}}{\partial q_{it}} = \frac{1}{\mu_{it}}. \tag{9}$$

Indeed, choosing unit we can always define $\bar{\lambda}_{it} = b\lambda_{it} + c$ so that $\frac{\partial r_{it}}{\partial \bar{\lambda}_{it}} = \frac{1}{\mu_{it}}$. Therefore, imposing (9) is consistent with the "true" underlying λ_{it} being a linear transformation of what we are able to estimate: $\bar{\lambda}_{it}$. In what follows we thus simply refer to (9).

The key intuition behind (9) is that, under this assumption, everything works as if firms were selling quantities $\tilde{Q}_{it} = Q_{it}\Lambda_{it}$ while charging prices $\tilde{P}_{it} = P_{it}/\Lambda_{it}$ so generating revenues $R_{it} = P_{it}Q_{it} = \tilde{P}_{it}\tilde{Q}_{it}$. Yet, what are the properties of λ_{it} and how can it be formally interpreted? This is straightforward to answer in the case of monopolistic competition with a representative consumer. We show in Appendix A that monopolistic competition models with a representative consumer satisfy (9) as long as the consumption of a variety enters direct utility as $\tilde{Q}_{it}$ instead of Q_{it}. In light of this Λ_{it} is a measure of vertical differentiation or quality.[8] This means that any Melitz (2003)-type model featuring heterogeneity in productivity can be augmented with heterogeneity in demand across firms satisfying (9) irrespective of the underlying preferences: CES, Translog, CARA, etc. As far as oligopolistic competition models are concerned, the key intuition behind (9) is the same and we lay down in Appendix A an oligopolistic model based on Atkeson and Burstein (2008) to show how our framework fits into strategic interactions. We also provide in Appendix A an explicit framework, based on the discrete/continuous choice models class considered in Nocke and Schutz (2016), allowing to deal with our λ_{it} in random utility models.

Besides defining what we mean by product appeal, (9) is also important in our analysis for two reasons. First, it provides a simple way of quantifying demand heterogeneity. We can in fact write down, based on a first-order linear approximation, firm revenue as:[9]

$$r_{it} \approx \frac{1}{\mu_{it}}(q_{it} + \lambda_{it}), \tag{10}$$

from which λ_{it} can be computed by simply using data on quantity and revenue as well as an estimate of the markup μ_{it}:[10]

$$\lambda_{it} = \mu_{it} r_{it} - q_{it}. \tag{11}$$

[8]As discussed in Di Comite et al. (2014) clear definitions of horizontal and vertical differentiation until now only exist in discrete choice models with indivisible varieties and with consumers making mutually exclusive choices. Many discrete choice models actually incorporate both types of differentiation (Anderson et al., 1992). In contrast, a clear distinction between horizontal (taste) and vertical (quality) differentiation is to a great extent absent in models where consumers have a love for variety and purchase many products in different quantities. As in Di Comite et al. (2014) our approach is characterized by features including both horizontal and vertical differentiation, which Di Comite et al. (2014) refer to as "verti-zontal".

[9]Again, to simplify notation we ignore components that are constant across firms in a given time period.

[10]From now onwards we use $=$ rather than $\approx$ to alleviate notation.

Second, by allowing to explicitly write down revenue in (10) as a function of observables and heterogeneities, it provides us with an additional, to the quantity equation (1), source of identification for technology parameters. This is what we develop in the estimation strategy described in Section 3.

Two questions naturally arise at this stage. First, are markups in our framework constrained to be a residual dimension of heterogeneity in that they are completely determined by a firm's quantity TFP, product appeal and predetermined inputs? Second, can our framework go beyond the first-order linear approximation of the revenue equation provided by (10)? In order to address the first question and part of the second it is useful to consider the generalized CES preferences structure introduced by Spence (1976) within a monopolistic competition setting. In this specification a representative consumer demand is obtained from the following problem:

$$\max_{Q} \left\{ \int_{i \in I_t} \frac{\eta_{it}}{\eta_{it}-1} \left(\Lambda_{it} Q_{it}\right)^{\frac{\eta_{it}-1}{\eta_{it}}} \mathrm{d}i \right\} \text{ s.t. } \int_i P_{it} Q_{it} \mathrm{d}i = B_t,$$

where B_t is the budget, Q is a vector with elements Q_{it} and the set of varieties is denoted by I_t. Note that in this case each firm is characterized by two parameters governing demand: (i) a Λ_{it} satisfying (9); (ii) an elasticity of demand η_{it} that uniquely pins down the markup μ_{it}. The first order condition to the consumer problem implies:

$$P_{it} \kappa_t = \Lambda_{it}^{\frac{\eta_{it}-1}{\eta_{it}}} Q_{it}^{-\frac{1}{\eta_{it}}}, \tag{12}$$

where κ_t is a Lagrange multiplier. Re-arranging suggests that firm-level demand is:

$$Q_{it} = P_{it}^{-\eta_{it}} \Lambda_{it}^{\eta_{it}-1} \kappa_t^{-\eta_{it}}.$$

Profit maximization of firm i then requires:

$$P_{it} = \mu_{it} \frac{\partial C_{it}}{\partial Q_{it}},$$

where the markup of firm i is the usual function of the elasticity of demand: $\mu_{it} = \frac{\eta_{it}}{\eta_{it}-1}$. Despite simple and somewhat mechanical, the advantage of such a specification is to show that our framework is consistent with markups that are not fully determined by a_{it}, λ_{it}, k_{it} and l_{it}. Indeed, μ_{it} can in this case be arbitrarily correlated with those state variables. When turning to the data we will see later on that markups are indeed correlated with productivity, product appeal and predetermined inputs (capital stock in particular) in a way that makes sense. Yet, we will also see that a large proportion of markups variation across firms remains unexplained by these variables suggesting that demand heterogeneity might be richer than quality differences across products and we are able to capture this additional heterogeneity via μ_{it}. At the same time, we acknowledge this might well reflect model mispecifications and/or measurement error but we are not able with available data for this paper to draw the line between these two alternative explanations. We are currently pursuing this line of research with more suitable data for France.

Moving to the second question, in the monopolistic competition generalized CES case we have, by multiplying both sides of (12) by Q_{it} and taking logs while using $\mu_{it} = \frac{\eta_{it}}{\eta_{it}-1}$ and ignoring the constant κ_t:

$$r_{it} = \frac{1}{\mu_{it}} \left(q_{it} + \lambda_{it} \right), \tag{13}$$

meaning that the first-order linear approximation provided by (10) here holds as an equality. Therefore, if in a particular situation a generalized CES preferences structure with monopolistic competition is deemed to be rich enough to characterize demand heterogeneity and market structure, our framework can be used as an exact solution to this modeling choice. However, as far as estimation is concerned, the same econometric procedure developed in Section 3 applies to both (10) and (13). In the former case it builds upon a first-order linear approximation that is consistent with different hypotheses on preferences and market structure, while in the latter case it is the exact formula derived from generalized CES preferences of a representative consumer and monopolistic competition.

In order to further address the first-order linear approximation issue we provide here highlights of an alternative specification while relegating details to Appendix A. The key idea is the same as in the our main analysis, i.e., work out the math of the revenue equation. Specifically, we look at an additively separable utility function shaped like the Gaussian CDF:[11]

$$U(\tilde{q}) = \int_{i \in I_t} \Phi \left(\tilde{q}_{it}, \beta_0, \beta_1, \beta_2 \right) \mathrm{d}i$$

where $\tilde{q}_{it} = q_{it} + \lambda_{it}$ and $\Phi\left(\cdot\right)$ is the Gaussian cdf, i.e.,

$$\Phi \left(\tilde{q}_{it} \right) = u \left(\tilde{q}_{it} \right) = \int_{-\infty}^{\tilde{q}_{it}} \phi \left(\tilde{q}_{it} \right) \mathrm{d}\tau,$$

and we set $\phi\left(\tilde{q}_{it}\right) = \exp\left(-\beta_2^2 \tilde{q}_{it}^3 + \beta_1 \tilde{q}_{it} + \beta_0\right)$ but could have equally used more or less involved formulations. (9) holds here and so we could use the first-order linear approximation (10) for the revenue equation. However, we could also work out the algebra (as we do in Appendix A) to obtain the following exact formula for revenue equation:

$$r_{it} = \left(\frac{1}{3} \frac{1}{\mu_{it}} + \frac{2}{3} \beta_1 \right) \left(q_{it} + \lambda_{it} \right),$$

[11]See Berhold (1973) for further discussion of the Gaussian CDF as a utility function.

where the profit maximizing markup is a function of parameters and the optimal quantity produced: $\mu_{it} = (-3\beta_2^2\tilde{q}_{it}^2 + \beta_1)^{-1}$. We then show in Appendix A how to modify our estimation strategy to incorporate the exact formula and be able to ultimately recover quantity TFP, product appeal and markups.

2.4 Identification

Provided estimates of the technology parameters of the quantity equation (1), one can directly compute an estimate of quantity TFP ($\hat{a}_{it}$) and use, building on Hall (1986) result in (5), the elasticity of output with respect to materials to get an estimate of markups ($\hat{\mu}_{it}$) and subsequently use (11) to get an estimate of product appeal ($\hat{\lambda}_{it}$). Furthermore, we have seen in the previous Section that our framework in consistent with markups not being simply a residual dimension of heterogeneity, i.e., not necessarily fully determined by quantity TFP, product appeal and predetermined inputs. Therefore, the set of assumptions laid above is enough to separately identify our 3 sources of heterogeneity without imposing a priori any restrictions on their correlation.

The final step is thus to obtain estimates of the technology parameters. In the estimation procedure described in Section 3 we use both the revenue and quantity equations to back out technology parameters. In order to do so we build on the assumption that product appeal follows, like quantity TFP, a Markov process like in Jaumandreu and Yin (2017). Again, for ease of exposition, we consider here the leading AR(1) case while describing how to handle richer cases in Appendix C.[12] More specifically we assume:

$$\lambda_{it} = \phi_\lambda \lambda_{it-1} + \nu_{\lambda it} \tag{14}$$

where $\nu_{\lambda it}$ are iid and uncorrelated with past values of product appeal.[13] At the same time we allow ν_{ait} in (6) and $\nu_{\lambda it}$ in (14) to be correlated with each other. Note that, by allowing ν_{ait} and $\nu_{\lambda it}$ to be correlated with each other, we allow demand and productivity λ_{it} and a_{it} to be correlated with each other; something we will see later on as being a key feature of the analysis. In Appendix C we further consider other mechanisms leading to a correlation

[12]In Appendix C we consider more general cases including non-linear terms and unobserved time-invariant heterogeneity; the latter being something the standard proxy variable approach cannot handle. At the same time, we show in Appendix C how to extend our framework to endogenous processes based on a simple extension of the model developed in Doraszelski and Jaumandreu (2013). This simple extension highlights how to incorporate endogenous choices of the firm affecting productivity and/or product appeal while allowing both a_{it-1} and λ_{it-1} to be on the right hand side of equations (6) and (14). This is in addition to the extension of our model provided in Caliendo et al. (2015) and featuring the endogenous choice of the organizational structure affecting both revenue and physical productivity.

[13]More precisely, we posit that a_{it} and λ_{it} are jointly described by a VAR(1) process meaning that ν_{ait} and $\nu_{\lambda it}$ are both uncorrelated with past values of productivity and product appeal.

between λ_{it} and a_{it}.[14]

We believe (14) is a suitable assumption within our framework and provide later on evidence that a simple Markov process does indeed a good job in matching the time evolution of λ_{it}. Regarding the latter, we get autoregressive coefficients and R^2s for λ_{it} comparable to those of quantity TFP a_{it} and this is actually robust to controlling for fixed effects in (6) and (14). Regarding the former, we believe (14) is a reasonable assumption because λ_{it} captures consumers' overall evaluation of a firm's products quality and appeal; something that arguably does not change dramatically from one year to another. It takes years of effort and costly investments to firms to establish their brand and build their customers' base very much like it takes years of effort and costly investments to firms to put in place and develop an efficient production process for their products. We thus believe there are profound similarities, as well as synergies and interactions, between the processes of productivity and demand and so if the former can be approximated by a Markov process we do not see why the latter could not be. Of course, a Markov process for λ_{it} is, very much like the standard assumption in the productivity literature of a Markov process for a_{it}, mainly a convenient operational assumption used to obtain identification in the absence of more detailed information about the effort and costly investments, as well as about the synergies and interactions, we refer to in the above. In this respect we show in Appendix C how to extend our framework to endogenous processes based on a simple extension of the model developed in Doraszelski and Jaumandreu (2013).[15]

A complementary approach to ours, that we have experimented with using data in this paper as well as comparable production data for Brazil, France and the UK, would consist in not imposing a Markov process for λ_{it} while imposing, as in De Loecker et al. (2016), existence and invertibility of a suitable conditional input demand for materials. Under this approach, information coming from the revenue equation is not needed to recover technology parameters and the quantity equation can be estimated along the lines developed in De Loecker et al. (2016) to get estimates of technology parameters as well as quantity TFP ($\hat{a}_{it}$). One can then use the Hall (1986) result in (5) to get an estimate of markups ($\hat{\mu}_{it}$) as well as (11) to get an estimate of product appeal ($\hat{\lambda}_{it}$).

Results obtained using this complementary approach, that we label DLG-WLD and on which we will provide some highlights later on, are qualitatively, and to a large extent also quantitatively, identical to those we present below using (14) and the estimation procedure described in Section 3. However, we do believe our estimation approach is potentially superior

[14] We consider the presence of fixed effects in (6) and (14) as well as the presence of both a_{it-1} and λ_{it-1} on the right hand side of equations (6) and (14).

[15] This simple extension highlights how to incorporate endogenous choices of the firm affecting productivity and/or product appeal while allowing both a_{it-1} and λ_{it-1} to be on the right hand side of equations (6) and (14).

for the following reasons. First, existence and invertibility of a suitable conditional input demand for materials amount to make implicit (and unclear) assumptions about demand and market structure. For example, the conditional input demand for materials m_{it} that we derive in Appendix E depends upon the full set of heterogeneities and the initial capital stock (a_{it}, λ_{it}, μ_{it} and k_{it}). The proxy variables used to control for our λ_{it} and μ_{it} in De Loecker et al. (2016), namely prices and market shares, clearly also depend upon the full set of heterogeneities and the initial capital stock and it is not guaranteed that: (i) one can express materials' demand as a function of capital, quantity TFP, output price and market share; (ii) invertibility of this specific conditional input demand function with respect to a_{it} holds. Whether (i) and (ii) are satisfied ultimately depends upon some specific features of demand and market structure that are implicit in the problem. Besides, we show in Appendix E that invertibility of the conditional input demand does not generally apply even when specified in terms of a_{it}, λ_{it}, μ_{it} and k_{it} because of the presence of firms with low markups.

Second, in the estimation procedure described in De Loecker et al. (2016) firm lagged market share and price are, among other things, added as covariates in the quantity equation (1). In this respect note that using the market share of firm i as a proxy amounts to using the revenue of firm i as a proxy. Indeed, market share is firm revenue divided by industry-level sales and the denominator is constant across firms. Lagged revenue and price obviously perfectly predict lagged quantity which is a very powerful predictor of current quantity who is on the left-hand side of (1). Therefore, there is very little data variation left to identify technology parameters so leading to potentially noisy estimates of α_M, α_L and γ. This is made even more problematic by the need to actually use, in order to convincingly approximate the conditional input demand, a polynomial expansion of (lagged) revenue, price, materials, labour and capital as additional covariates in (1).[16]

2.5 Revenue productivity

Where our framework delivers its full potential and ultimately provides what we believe it is an important contribution is not just in getting the TFP "more right" than in other methodologies but rather in allowing to unravel many dimensions of heterogeneity potentially correlated with each other and quantify them. This in turns allows us to derive an exact decomposition of revenue productivity in terms of these heterogeneities so bridging the gap between quantity TFP estimations and revenue TFP estimations.

Revenue TFP is defined as $TFP^R_{it} \equiv r_{it} - \bar{q}_{it}$ where $\bar{q}_{it}$ is an index of inputs use that we

[16] As matter of fact, after experimenting several possibilities with data for different countries (Belgium, Brazil, France and the UK) we ended up dropping revenue from the control function and using only firm price as a proxy as originally suggested in De Loecker and Goldberg (2014). We found this choice to produce much more reliable production function coefficients within the De Loecker et al. (2016) quantity TFP estimation.

label scale: $\bar{q}_{it} \equiv q_{it} - a_{it} = \alpha_L l_{it} + \alpha_M m_{it} + (\gamma - \alpha_L - \alpha_M) k_{it}$. By using equation (10) and substituting we get:

$$TFP^R_{it} = \frac{1}{\mu_{it}} (a_{it} + \lambda_{it}) + \frac{1 - \mu_{it}}{\mu_{it}} \bar{q}_{it}, \tag{15}$$

meaning that TFP^R_{it} is a (non-linear) function of a, λ, μ and production scale. (15) can also be made linear by considering markups-adjusted quantity TFP, product appeal and scale: $\tilde{a}_{it} = \frac{a_{it}}{\mu_{it}}$, $\tilde{\lambda}_{it} = \frac{\lambda_{it}}{\mu_{it}}$, $\tilde{\bar{q}}_{it} = \frac{(1-\mu_{it})\bar{q}_{it}}{\mu_{it}}$:

$$TFP^R_{it} = \tilde{a}_{it} + \tilde{\lambda}_{it} + \tilde{\bar{q}}_{it}. \tag{16}$$

Equations (15) and (16) allow disentangling differences across firms and/or changes across time of revenue TFP into underlying variation/changes in quantity TFP, product appeal, markups and production scale. This in turns enables gaining deeper and sharper insights into productivity questions. For example, we show later on in Section 7.2 how changes in firm revenue productivity spurred by import competition from China materialize as the outcome of complex changes in quantity TFP, product appeal, markups and production scale. This in turns allows to better understand firm behaviour and margins of adjustment under competitive pressure and learn important lessons that can applied to other contexts. By decomposing the revenue productivity advantage of importers, we further provide in Section 7.3 evidence that demand heterogeneity across firms is at least as important as quantity TFP in drawing the line between importing and non-importing firms. This suggests that international trade model à la Melitz (2003) or Antras and Helpman (2004) should devote more attention to demand heterogeneity while at the same time trade policies should be paying equal attention to production efficiency and quality/marketing issues. In a related project (Jacob and Mion, 2017), we instead dissect the revenue productivity advantage of large cities for France and show it ultimately boils down to differences in production scale across space rather than differences in quantity TFP and/or product appeal so questioning standard models of agglomeration economies.

3 Estimation strategy

Our estimation procedure builds upon (10) and uses both the revenue and quantity equations to estimate technology parameters. We describe the procedure in Section 3.2 while in the next Section we explain how we deal with measurement error in output. Section 3.3 elaborates on what can be estimated and how when data on quantity is not available.

3.1 Measurement error in output

One issue we need to first account for is the presence of measurement error in quantity and/or revenue. In the case of quantity, instead of q_{it}, the econometrician might be observing q'_{it}=$q_{it} + e_{it}$ where e_{it} is measurement error. (1) thus becomes:

$$q'_{it} = \alpha_L l_{it} + \alpha_M m_{it} + (\gamma - \alpha_L - \alpha_M)k_{it} + a_{it} + e_{it}.$$

The approach suggested by the literature (Ackerberg et al., 2015; De Loecker et al., 2016) to deal with measurement error e_{it} is based on the proxy variable framework and a semi-parametric implementation. We follow this approach and estimate:

$$q'_{it} = poly(l_{it}, m_{it}, p_{it}, k_{it}) + e_{it}. \tag{17}$$

where q'_{it} is (log) quantity as reported in the data and $poly(.)$ is a third-order polynomial in l_{it}, m_{it}, p_{it} and k_{it}. We then use the OLS prediction of q'_{it} ($\hat{q'}^{OLS}_{it}$) as quantity in the rest of the analysis. We also use the same approach for revenue and run a regression similar to (17) to get $\hat{r}^{OLS}_{it}$. For both quantity and revenue we augment regressions with a full battery of 8-digit product dummies as well as year dummies. Our key findings are unaffected by accounting for measurement error with this approach but the reliability and precision of technology parameters does benefit from it.

3.2 Estimation Procedure

Our estimation procedure builds upon (10) and uses both the revenue and quantity equations to estimate technology parameters. The two-steps procedure described below is not the only one that can be used to recover technology parameters under our set of assumptions but has the advantage of being simple to implement and linear. By using more systematically parameters' constraints in the two equation and performing a joint, rather than two-steps, estimation one can possibly get more precise estimates. However, this would come at the cost of using a non-linear GMM estimator with non-linear parameters' constraints which is both more complex and numerically less well-behaved than the linear estimators we use.

By substituting q_{it} with the formula of the Cobb-Douglas we can transform (10) further as:

$$r_{it} = \frac{\alpha_L}{\mu_{it}} (l_{it} - k_{it}) + \frac{\alpha_M}{\mu_{it}} (m_{it} - k_{it}) + \frac{\gamma}{\mu_{it}} k_{it} + \frac{1}{\mu_{it}} (a_{it} + \lambda_{it}).$$

Furthermore, by using (5), we get:

$$LHS_{it} \equiv \frac{r_{it} - s_{Mit} (m_{it} - k_{it})}{s_{Mit}} = \frac{\alpha_L}{\alpha_M} (l_{it} - k_{it}) + \frac{\gamma}{\alpha_M} k_{it} + \frac{1}{\alpha_M} (a_{it} + \lambda_{it}). \tag{18}$$

where LHS_{it} is made out of observables only.

We then build upon our assumptions on the time process for a_{it} and λ_{it}: (6) and (14). However, before substituting (6) and (14) into (18) we need to find a convenient way to express a_{it-1} and λ_{it-1}. By using (5) and (10) we have:

$$\lambda_{it-1} = r_{it-1}\mu_{it-1} - q_{it-1} = r_{it-1}\frac{\alpha_M}{s_{Mit-1}} - q_{it-1}. \tag{19}$$

At the same time plugging (19) into (18) and re-arranging yields:

$$a_{it-1} = \alpha_M LHS_{it-1} - \alpha_L\left(l_{it} - k_{it}\right) - \gamma k_{it-1} - \left(r_{it-1}\frac{\alpha_M}{s_{Mit-1}} - q_{it-1}\right). \tag{20}$$

Finally, by combining (6), (14), (19) and (20) into (18) we obtain:

$$\begin{aligned} LHS_{it} &= \frac{\gamma}{\alpha_M}k_{it} + \frac{\alpha_L}{\alpha_M}\left(l_{it} - k_{it}\right) + \phi_a LHS_{it-1} - \phi_a\frac{\gamma}{\alpha_M}k_{it-1} - \phi_a\frac{\alpha_L}{\alpha_M}\left(l_{it-1} - k_{it-1}\right) \\ &+ \left(\phi_\lambda - \phi_a\right)\left(\frac{r_{it-1}}{s_{Mit-1}} - \frac{q_{it-1}}{\alpha_M}\right) + \frac{1}{\alpha_M}\left(\nu_{ait} + \nu_{\lambda it}\right). \end{aligned} \tag{21}$$

Note that the revenue equation (21) is, besides the idiosyncratic productivity and demand shocks ν_{ait} and $\nu_{\lambda it}$, now entirely written in terms of observables and useful parameters. There are various ways of estimating (21) and here we use perhaps the simplest one. More specifically, we rewrite (21) as the following linear regression:

$$LHS_{it} = b_1 z_{1it} + b_2 z_{2it} + b_3 z_{3it} + b_4 z_{4it} + b_5 z_{5it} + b_6 z_{6it} + b_7 z_{7it} + u_{it}, \tag{22}$$

where z_{1it}=k_{it}, z_{2it}=$(l_{it} - k_{it})$, z_{3it}=LHS_{it-1}, z_{4it}=k_{it-1}, z_{5it}=$(l_{it-1} - k_{it-1})$, z_{6it}=$\frac{r_{it-1}}{s_{Mit-1}}$, z_{7it}=q_{it-1}, u_{it}=$\frac{1}{\alpha_M}\left(\nu_{ait} + \nu_{\lambda it}\right)$ as well as b_1=$\frac{\gamma}{\alpha_M}$, b_2=$\frac{\alpha_L}{\alpha_M}$, b_3=ϕ_a, b_4=$-\phi_a\frac{\gamma}{\alpha_M}$, b_5=$-\phi_a\frac{\alpha_L}{\alpha_M}$, b_6=$(\phi_\lambda - \phi_a)$, b_7=$-\left(\phi_\lambda - \phi_a\right)\frac{1}{\alpha_M}$. Given our assumptions, the error term u_{it} in (22) is uncorrelated with current capital and labour as well as with lagged inputs use, quantity and revenue.[17] Therefore, z_{1it} to z_{7it} are uncorrelated to u_{it} and (22) can be estimated by OLS. After doing this we set $\widehat{\frac{\gamma}{\alpha_M}}$=$\hat{b}_1$, $\widehat{\frac{\alpha_L}{\alpha_M}}$=$\hat{b}_2$ and $\hat{\phi}_a$=$\hat{b}_3$ and do not exploit parameters' constraints in the estimation.[18]

[17] If one allows labour to be a semi-flexible input then z_{2it} will be endogenous here. Yet, the lagged value $(l_{it-2} - k_{it-2})$ can, among others, be used as an instrument and (22) can be estimated via linear IV.

[18] This means that, for example, we do not exploit the non-linear constraints b_4=-b_1b_3 and b_5=-b_2b_3. We can certainly do this at the cost of using non-linear OLS. Furthermore, by exploiting parameters' constraints we could actually also estimate α_L and α_M, and so γ, from (22) without need for further estimations. However, a closer inspection to (22) reveals that identification of α_L, α_M and γ 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. By complementing the estimation with a second stage quantity equation we avoid these issues. Also note that $\phi_\lambda - \phi_a \simeq 0$ does not represent an issue for separately identifying productivity and product appeal when information on both quantity and revenue is available because productivity and

We now turn to estimating γ from the quantity equation in a second step. Combining (1) and (5) we have:

$$q_{it} = \mu_{it} s_{Mit} \left(m_{it} - k_{it}\right) + \alpha_L \left(l_{it} - k_{it}\right) + \gamma k_{it} + a_{it}. \tag{23}$$

Further using $\alpha_M = \frac{\gamma}{b_1}$ as well as $\alpha_L = \frac{\gamma b_2}{b_1}$ and we get:

$$q_{it} = \frac{\gamma}{\hat{b}_1} \left(m_{it} - k_{it}\right) + \frac{\gamma \hat{b}_2}{\hat{b}_1} \left(l_{it} - k_{it}\right) + \gamma k_{it} + a_{it}, \tag{24}$$

where we replace b_1 and b_2 with their estimates $\hat{b}_1$ and $\hat{b}_2$ coming from (22). Finally, using (6) to substitute for a_{it} and using (20) we obtain:

$$\begin{aligned} q_{it} &= \frac{\gamma}{\hat{b}_1} \left(m_{it} - k_{it}\right) + \frac{\gamma \hat{b}_2}{\hat{b}_1} \left(l_{it} - k_{it}\right) + \gamma k_{it} + \gamma \frac{\hat{\phi}_a}{\hat{b}_1} LHS_{it-1} - \frac{\gamma \hat{b}_2 \hat{\phi}_a}{\hat{b}_1} \left(l_{it-1} - k_{it-1}\right) \\ &- \gamma \hat{\phi}_a k_{it-1} - \hat{\phi}_a \left(r_{it-1} \frac{\gamma}{\hat{b}_1 s_{Mit-1}} - q_{it-1} \right) + \nu_{ait}. \end{aligned} \tag{25}$$

Note that the only unobservable in (25) is the idiosyncratic productivity shock ν_{ait} while the only parameter left to identify is γ. We can more compactly write (25) as the following linear regression:

$$\overline{LHS}_{it} = b_8 z_{8it} + \nu_{ait} \tag{26}$$

where:

$$\begin{aligned} \overline{LHS}_{it} &= q_{it} - \hat{\phi}_a q_{it-1} \\ z_{8it} &= \frac{1}{\hat{b}_1} \left(m_{it} - k_{it}\right) + \frac{\hat{b}_2}{\hat{b}_1} \left(l_{it} - k_{it}\right) + k_{it} + \frac{\hat{\phi}_a}{\hat{b}_1} LHS_{it-1} \\ &- \frac{\hat{b}_2 \hat{\phi}_a}{\hat{b}_1} \left(l_{it-1} - k_{it-1}\right) - \hat{\phi}_a k_{it-1} - \frac{\hat{\phi}_a r_{it-1}}{\hat{b}_1 s_{Mit-1}} \end{aligned}$$

as well as $b_8 = \gamma$. Concerning z_{8it} we can use several moment conditions for identification: $E\left\{\nu_{ait} k_{it}\right\} = E\left\{\nu_{ait} l_{it}\right\} = E\left\{\nu_{ait} l_{it-1}\right\} = E\left\{\nu_{ait} m_{it-1}\right\} = E\left\{\nu_{ait} k_{it-1}\right\} = E\left\{\nu_{ait} q_{it-1}\right\} = E\left\{\nu_{ait} r_{it-1}\right\} = 0$.[19] IV estimation of (26) provides an estimate of γ that, together with $\widehat{\frac{\gamma}{\alpha_M}}$ and $\widehat{\frac{\alpha_L}{\alpha_M}}$ coming from the first stage revenue equation, uniquely delivers production function parameters ($\hat{\alpha}_L$, $\hat{\alpha}_M$ and $\hat{\gamma}$). We can then use (1), (5) and (11) to get estimates of quantity

product appeal are obtained from two different equations, (1) and (11), that are not affected by $\phi_\lambda - \phi_a \simeq 0$. We show later on how $\phi_\lambda - \phi_a \simeq 0$ actually allows recovering a combined $a_{it} + \lambda_{it}$ measure when only revenue data is available.

[19] If one allows labour to be a semi-flexible input then $E\left\{\nu_{ait} l_{it}\right\} = 0$ will not hold. However, all of the other moment conditions will hold and there are plenty to choose from.

TFP, markups and product appeal. Finally, we augment both the first and second-stage regressions with a full battery of 8-digit product dummies as well as year dummies. Standard errors are obtained via bootstrapping by re-sampling residuals in regressions (22) and (26).

3.3 What if quantity data is not available?

In many applications quantity data is simply not available to researchers and revenue is used instead as a left-hand side variable. In such cases, unless one is willing to make the strong assumption of constant and common markups across firms, like in Klette and Griliches (1996) and De Loecker (2011), it is not clear what revenue productivity estimators actually measure if anything at all. This is known in the literature as the output price bias (Klette and Griliches, 1996).

More specifically, prices would add to productivity shocks on the right-hand side of the quantity equation leading to $r_{it} = \alpha_L l_{it} + \alpha_M m_{it} + (\gamma - \alpha_L - \alpha_M)k_{it} + a'_{it}$, where $a'_{it} = a_{it} + p_{it}$ is the relevant error term. Even in a world where the only heterogeneity across firms is productivity, recovering technology parameters from the above revenue equation is problematic. Consider, for example, capital. The correlation of k_{it} with productivity a_{it} can be accounted for within the proxy-variable approach by assuming that capital is predetermined while a_{it} follows a Markov process and invertibility applies. Yet the correlation of capital with prices, the other component of the error term a'_{it}, is troublesome. The assumption of a Markov process for prices would be too hard to swallow. At the same time, even if capital is predetermined, its stock will affect short-term marginal costs and so prices in t. In this respect, instruments would not be of any help here because one would need instruments that are correlated with k_{it} but that are not correlated with prices that depend upon k_{it}. Similar arguments can be made about materials and labour with details depending upon the specific assumptions made about those inputs.

Our framework can be applied to situations where quantity data is not available and allows recovering a composite measure of productivity and product appeal ($\omega_{it} = a_{it} + \lambda_{it}$) as well as markups ($\mu_{it}$). We label this variant of our model MUOMEGA. This is straightforward to implement and simply requires making a simple additional assumption ($\phi_a \simeq \phi_\lambda$); something for which we find later on support in the data.

To the extent that roughly the same degree of persistency characterizes the productivity and product appeal processes we have $\phi_\lambda - \phi_a \simeq 0$ and so (21) simplifies to:

$$LHS_{it} = \frac{\gamma}{\alpha_M} k_{it} + \frac{\alpha_L}{\alpha_M}(l_{it} - k_{it}) + \phi_a LHS_{it-1} - \phi_a \frac{\gamma}{\alpha_M} k_{it-1} - \phi_a \frac{\alpha_L}{\alpha_M}(l_{it-1} - k_{it-1}) \frac{1}{\alpha_M}(\nu_{ait} + \nu_{\lambda it}). \tag{27}$$

Equation (27) can be estimated by OLS and allows recovering estimates of the reduced form parameters $\frac{\gamma}{\alpha_M}$ and $\frac{\alpha_L}{\alpha_M}$. With these at hand one can use (18) to get:

$$\frac{\widehat{(a_{it}+\lambda_{it})}}{\alpha_M} = LHS_{it} - \frac{\widehat{\gamma}}{\alpha_M}k_{it} - \frac{\widehat{\alpha_L}}{\alpha_M}\left(l_{it}-k_{it}\right). \tag{28}$$

Equation (28) thus enables recovering (up to the innocuous scaling factor α_M) a composite measure of productivity and product appeal $(a_{it}+\lambda_{it})$ while allowing for heterogeneous markups (μ_{it}). As far as the latter are concerned, they can also be identified up to a scale using the estimate $\widehat{\frac{\gamma}{\alpha_M}}$. More specifically, from (5) we have:

$$\frac{\widehat{\mu_{it}}}{\gamma} = \frac{\frac{\widehat{\alpha_M}}{\gamma}}{s_{Mi}}. \tag{29}$$

Further note that, as long as returns to scale are approximately constant ($\gamma \simeq 1$), which we find to be largely the case for all the industries we study below, one gets $\frac{\widehat{\mu_{it}}}{\gamma} \simeq \widehat{\mu_{it}}$. Furthermore, in this instance also revenue productivity (15) can be recovered up to a the scaling factor α_M:

$$\frac{\widehat{TFP^R_{it}}}{\alpha_M} = \frac{1}{\widehat{\mu_{it}}}\frac{\widehat{(a_{it}+\lambda_{it})}}{\alpha_M} + \frac{1-\widehat{\mu_{it}}}{\widehat{\mu_{it}}}\frac{\widehat{\bar{q}_{it}}}{\alpha_M}, \tag{30}$$

where $\frac{\widehat{\bar{q}_{it}}}{\alpha_M} = \frac{\widehat{\alpha_L}}{\alpha_M}\left(l_{it}-k_{it}\right) + \left(m_{it}-k_{it}\right) + \frac{\widehat{\gamma}}{\alpha_M}k_{it}$. This version of the model can thus be implemented to those frequent situations where only revenue data is available. In this respect, Doraszelski and Jaumandreu (2013) and Jaumandreu and Yin (2017) show how to get even more from revenue data if additional information on firm-level price changes (Doraszelski and Jaumandreu, 2013) or sales across different markets (Jaumandreu and Yin, 2017) is available.

4 Extensions

In Appendices A to D we show how to interpret and apply the key property (9) to various demand and market structure settings as well as how to extend our analysis to a wider set of production functions, to richer processes for productivity and product appeal and to multi-product firms. In Appendix E we derive the conditional input demand for materials and show how invertibility does not generally hold.

In Appendix A we show that monopolistic competition models with a representative consumer satisfy (9) as long as quantities enter direct utility as $\tilde{Q}_{it} = \Lambda_{it}Q_{it}$. Therefore, for any preferences structure that can be used to model monopolistic competition and that can be described by a well-behaved differentiable direct utility we can, starting from the baseline formulation $U(Q)$, introduce product appeal in such a way that (9) is satisfied. This includes of course the CES and its generalizations as well as the CARA preferences used in Behrens et al. (2014) and the Translog preferences featuring in Feenstra (2003) and Rodríguez-López (2011). We then move beyond representative consumer models and consider discrete/continuous choice models. Discrete/continuous choice models represent a generalisation of standard discrete choice models including, for example, the Multinomial Logit. They are obtained from a random utility framework in which consumers not only choose one alternative amongst many but also how much to consume of a particular good. We consider the class of models covered in Nocke and Schutz (2016) and show how (9) emerges by an appropriate choice of the indirect utility.

We subsequently develop in Appendix A an oligopoly model based on Atkeson and Burstein (2008) and further refined in Hottman et al. (2016) for multi-product firms. The key ingredient to get (9) is the same as in the monopolistic competition case, namely that quantities enter into preferences as $\tilde{Q}_{it} = \Lambda_{it} Q_{it}$. Atkeson and Burstein (2008) consider a nested CES model of quantity competition à la Cournot in which firms sell differentiated varieties and are large enough to perceive their impact on industry aggregates while charging markups that depend upon their market share. We introduce product appeal in their model and show how (9) works. We also lay down the key ingredients needed to extend the analysis to multi-product firms based on Hottman et al. (2016). Finally, we show how our model can be extended beyond (9), i.e., without resorting to any linear approximation by fully specifying a preference structure and working out the corresponding algebra for the revenue equation. We do so in the case of an additively separable utility function shaped like the Gaussian CDF.

Appendix B is devoted to extending the analysis to more flexible production functions. In particular, we consider the (homogenous) translog form and show how to modify the estimation procedure accordingly. In Appendix C we instead provide a number of examples of richer processes for a and λ that can be dealt with. More specifically, we consider: (i) a non-linear Markov process for a and λ; (ii) the presence of time-invariant unobserved heterogeneity. We show the former case complicates the algebra but not the underlying structure of the problem while the latter case can be accommodated quite easily. This is in stark contrast with the standard proxy variable approach that cannot handle the presence of correlated time-invariant heterogeneity. We also explain how to handle measurement error in capital within our framework. Moving forward, we show how to extend our framework to endogenous processes based on a simple extension of the model developed in Doraszelski and Jaumandreu (2013). This simple extension highlights how to incorporate endogenous choices of the firm affecting productivity and/or demand while allowing both a_{it-1} and λ_{it-1} to be on the right hand side of equations (6) and (14). This is in addition to the extension of our model provided in Caliendo et al. (2015) and featuring the endogenous choice of the organizational structure affecting both revenue and physical productivity.

We consider multi-product firms in Appendix D. There are several issues related to multi-product firms. We focus on the issue of the assignment of inputs to outputs. Produced quantities and generated revenues may be observable for the different products of each firm in databases like ours. However, information on inputs used for a specific product is typically not available. We propose in Appendix D an extension of our baseline model to solve the problem of assigning inputs to outputs for multi-product firms. In doing so we assume, as in De Loecker et al. (2016), there is a limited role for economies (or diseconomies) of scope on the cost side. However, contrary to De Loecker et al. (2016), we do not impose multi-product firms to be characterized by a common productivity across the different products they produce. We also allow for firm-product-time specific markups but impose product appeal/quality to be common across products within a firm. This corresponds to a setting where firms can be distinguished into those consistently selling high quality products and those consistently selling low quality products. Yet firms are allowed to be more or less efficient in the production of a specific product and charge different markups. The assumptions we lay down and the related estimation procedure are consistent with both a monopolistically competitive market structure, like the one developed in Bernard et al. (2011), and the oligopoly model developed in Hottman et al. (2016) that we discuss in Appendix A.

Finally, we derive in Appendix E the conditional input demand for materials as a function of a_{it}, λ_{it}, μ_{it}, and k_{it} and show that invertibility with respect to a_{it} does not generally apply because of the presence of firms with low markups: μ_{it} smaller than $\alpha_L + \alpha_M$. Such firms are present in our data as well as in other datasets and analyses. For example, given an average markup in between 1.1 and 1.2 and a standard deviation of 0.5, quite a few firms in De Loecker and Warzynski (2012) would fall into this category. Yet, this is not an issue within our estimation procedure because invertibility is not needed.

5 Data, descriptives and additional variables

5.1 Basic 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 whose first four digits come from NACE codes). 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). During our sample period, each Belgian firm with 10 employees or more - or with a revenue greater than a certain threshold in a given year - had to fill out the survey.[20] 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 6,000 firms a year over the period of 1995 to 2007. 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 as well as by De Loecker et al. (2014) for their study of the links between international competition and firm performance. There are about 4,500 distinct 8-digit products within the Prodcom classification. The level of detail is such that, for example, we are able within the "Meat and Meat products" industry (NACE code 151) to look at specific products like "Sausages not of liver" (Prodcom code 15131215) and "Fresh or chilled cuts of geese; ducks and guinea fowls" (Prodcom code 15121157) while within the NACE 212 "Article of Papers" industry, we can distinguish between products like "Envelopes of paper or paperboard" (Prodcom code 21231230) and "Wallpaper and other wall coverings; window transparencies of paper" (Prodcom code 21241190). Despite not quite as detailed as product categories available with bar-code data for retail products, our dataset has the advantage of spanning across the entire manufacturing sector.

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.

Besides, 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 import 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.

5.2 Additional variables and descriptives

In the analysis of the impact of import competition from China on revenue productivity and its components, that we report in Section 7.2, we further use additional trade and import

[20] Rules are somewhat different for other EU countries. In particular some EU countries only surveyed firms with 20 or more employees. The 10 employees threshold has been recently increased to 20 in Belgium as well.

quota data. The trade data comes from the Comtrade database provided by the United Nations. We use EU-15 and US imports data over the period 1995 to 2007 at the HS6-digit level to construct a measure of Chinese imports penetration in these two markets. We first build on a concordance between the HS6-digit classification and the CPA6-digit classification, where the latter dictates the first 6 digits of Prodcom codes, to measure imports at the CPA6-exporting country-year level.[21] We then construct the following measure of Chinese imports penetration, in either the EU15 or the US market, for each CPA6-digit product and time t:

$$IPC^{mkt}_{CPA6,t} = \frac{IMP^{mkt}_{CPA6,China,t}}{\sum_c IMP^{mkt}_{CPA6,c,t}}. \tag{31}$$

where $IMP^{mkt}_{CPA6,c,t}$ are imports by market $mkt = \{EU15, US\}$ of products belonging to a specific $CPA6$ product from country c at time t while $IMP^{mkt}_{CPA6,China,t}$ represents imports from China by either the EU15 or the US of products belonging to a specific $CPA6$ product at time t. A similar measured has been used in Mion and Zhu (2013) for Belgium, as well as by a number studies for other countries, for the analysis of the various economic impacts of import competition from China. In this respect, we believe the relevant market for the highly export-oriented Belgian firms over our time frame has to be considered the EU15. Therefore, $IPC^{EU15}_{CPA6,t}$ is our preferred measure of the import competition faced by Belgian firms producing products belonging to a specific $CPA6$ code at time t. At the same time we allow, as in Autor et al. (2013), for the presence of unobserved demand/technology shocks at the $CPA6$-time level characterizing the EU15 market, and correlated with $IPC^{EU15}_{CPA6,t}$, by instrumenting $IPC^{EU15}_{CPA6,t}$ with $IPC^{US}_{CPA6,t}$. More specifically, we match in each year firms to our $CPA6$-time measure[22] and regress firm revenue productivity on $IPC^{EU15}_{CPA6,t}$. While allowing for firm fixed effects and time dummies, we subsequently instrument $IPC^{EU15}_{CPA6,t}$ with $IPC^{US}_{CPA6,t}$. We then do the same for the different components of revenue productivity (a_{it}, λ_{it}, etc.).

As a complementary attempt to identify the impact of import competition from China on revenue productivity and its components we focus on a specific industry ("Textile and Apparel") and exploit detailed HS6-level information on import quotas. These quotas were imposed at the EU-level on Chinese imports, as well as on imports from other non-WTO countries, and affected some products within the industry but not others. As a consequence

[21] The concordance between HS6 and CPA6 is quite straightforward and we have used suitable tables provided by the RAMON EU website. The same does not apply to the concordance between HS6 and Prodcom 8-digit. This is the reason why we have decided to work at the CPA6 level. The CPA 6-digit still represents a very detailed breakdown of products. For example, there are 1,370 distinct CPA6 products corresponding to about 4,500 Prodcom 8-digit products.

[22] As explained below we focus in our analysis on firm producing a single Prodcom 8-digit product. Therefore, we match firms to $IPC^{EU15}_{CPA6,t}$ based on the first 6-digits of the unique Prodcom 8-digit product produced by firm i at time t.

of China joining the WTO, these quotas were removed over our time frame. To provide some context, when these quotas were abolished this generated a 240% increase in Chinese imports on average within the affected product groups. The data and estimation strategy are borrowed from Bloom et al. (2016) to which the reader may refer for further details. The underlying identifying assumption of this strategy is that unobserved demand/technology shocks are uncorrelated with the strength of quotas to non-WTO countries (like China) in 2000. Since these quotas were built up from the 1950s, and their phased abolition negotiated in the late 1980s was in preparation for the Uruguay Round, Bloom et al. (2016) conclude that this seems like a plausible assumption. Operationally we compute, for each 6-digit CPA product category, the proportion of 6-digit HS products that were covered by a quota, weighting each HS6 product by its share of EU15 imports value computed over the period 1995-1997. We label this $QUOTA_{CPA6}$ and focus on the period 1998-2007 to analyze firm behavior. More specifically, we match in each year firms to $QUOTA_{CPA6}$ and run a simple diff-in-diff specification where the time-change in, for example, firm revenue productivity is regressed on $QUOTA_{CPA6}$. We then do the same for the different components of revenue productivity.

With the exception of the quota analysis where we consider the sub-sample 1998-2007, we focus our study on the period 1996-2007 for which all datasets are available and during which there has not been any major change in data collection and data nomenclatures (such as the NACE and the CN nomenclatures changing considerably in 2008).[23]

We choose not to analyse multi-product firms in this paper and focus on single-product firms, i.e., firms producing a single 8-digit Prodcom product. This is for a variety of reasons. First, we could definitely include multi-product firms in our analysis by drawing on the extension of the MULAMA model presented in Appendix D. Yet, in doing so we would not be able to tell whether our findings are specific to the MULAMA model or whether they are driven by the additional assumptions we need to make in order to solve the inputs assignment problem for multi-product firms. In this respect, we believe focusing on single-product firms allows for a more direct comparison to other approaches and so better highlights our methodological contribution. Second, in order to estimate the MULAMA model on multi-product firms we first need, as in De Loecker et al. (2016), to restrict to the single-product firms sample to

[23]As reported in, for example, Bernard et al. (2012b) there have been some minor changes in 8-digit level Prodcom codes during our sample period. More specifically, the first 6-digits of Prodcom codes have remained virtually unchanged from 1996 to 2007 because they correspond to the CPA classification and this has barely changed over our sample period. Therefore, most code changes involved the last two-digits of Prodcom product codes. However, whenever a new code is introduced the old code is not re-assigned any more. Rather than attempting a complicated, and potentially imprecise, Prodcom 8-digit time concordance exercise we have decided to use the original codes. This means that, if a product changes code in year t, we will be using two dummies in the estimations; one for the code prior to t and one for the code from t. The same principle applies to one-to-many, many-to-one and many-to-many codes changes. Albeit conservative, we believe this is the best solution in this case.

estimate the parameters of the production function and there actually plenty such firms (more than half of firm-year pairs) in the data. Third, in focusing on single-product firms we improve upon previous analyses of the impact of Chinese imports competition by being able to very precisely match a measure of import competition with what the firm actually produces. For example, Mion and Zhu (2013) and Bloom et al. (2016) combine information on the primary 4-digits industry code of a firm with some of the available secondary codes to account for the fact that some firms are active in many industries while the measure they use is specific to an industry. In this respect our information is more detailed, 6-digits CPA, while at the same time we are fully confident that the CPA-specific measure of Chinese import competition we attach to a firm covers all of its products.

As in previous studies using either revenue or quantity data our estimations are run at a more aggregate level, that we label "industry" g, rather than at the finest available classification (8-digits products). This is needed to have a sufficiently large number of observations to estimate production function parameters in a consistent way. Operationally, this means we estimate (22) and (26) separately for each industry by pooling together firms producing 8-digits products belonging to a given industry while at the same time adding time and 8-digits product dummies. This amounts to assume that technology parameters are the same across products within an industry. Yet, we use actual quantities (and revenues) corresponding to the specific 8-digits product produced by a firm.

In terms of data cleaning, besides getting rid of missing and/or inconsistent observations, we exclude from the analysis firms that in a given year report different sales (± 15%) in the Prodcom and the Balance sheet data. Indeed, for some firms manufacturing is only one part of their operations. We then apply a 1% top and bottom trimming based on the following variables: (i) value added over revenue; (ii) materials expenditure over revenue; (iii) capital stock over labour expenditure; (iv) price within an 8-digit product.

Insert Table 1 and Figure 1 about here.

Table 1 provides our industry breakdown as well as some basic summary statistics (mean, standard deviation, 5th and 95th percentiles) and the number of observations for the estimation sample. At the same time Figure 1 plots, after demeaning for each 8-digit product, the log quantity and log price corresponding to each firm-year pair in our sample. We provide a plot for each industry and, as one can appreciate, there is indeed a negative correlation (within 8-digit products) between prices and quantities. However, the correlation is far from perfect with many instances of firms selling substantially higher quantities than others for the same price and vice versa. In this respect differences are indeed substantial because prices and quantities in Figure 1 are in log units. Overall, this points to a fair amount of heterogeneity in demand across firms in the data.

6 Estimation results

In this Section we provide a number of summary statistics about our estimations, our measures of TFP, product appeal and markups and examine how the three dimensions of heterogeneity correlate with each other in a cross section as well as across time. We also briefly touch upon comparison to other methodologies.

6.1 Main results

Table 2 provides estimates of production function parameters for each industry obtained using our econometric procedure. Coefficients are in line with expectations and there seems to be overall support for constant returns to scale ($\gamma = 1$). Capital coefficients are sometimes small as usual, likely due to measurement error plaguing this variable, but they do not look any worse than those reported in De Loecker et al. (2016) using quantity data for India or to those we obtain using the DLG-WLD procedure on our data (see Table F-1 in Appendix F). Last but not least, the correlation between quantity TFP obtained with our procedure and quantity TFP obtained with the DLG-WLD procedure is 0.998 across all observations and 0.990 once demeaning both TFP measures by 8-digit product codes.

Insert Table 2 about here.

In terms of markups the average across all observations is 1.091 which is in line with numbers reported in, for example, De Loecker and Warzynski (2012). We further provide the density distribution of markups across observations separately for each industry in Figure 2 along with the corresponding mean (red vertical line). In this respect, Figure 2 points out how markups vary considerably across firms within each industry. As far as product appeal is concerned averages are, like for the case of quantity TFP, of little value. What it is interesting is variation in the data and Table 3 reports the standard deviation of 8-digit product code demeaned values of a_{it} and λ_{it} as well as raw values of μ_{it}. The key finding stemming from Table 3 is that, within products, there is as much variation in product appeal as there is variation in quantity TFP so confirming the first-hand impression stemming from raw data on prices and quantities plotted in Figure 1. Overall, this suggests that heterogeneity in demand is a key component of firm idiosyncracies being at least as sizeable as heterogeneity in productivity and so a potentially powerful key to unlock patterns in the data. Concerning markups they instead display considerably less variation than a_{it} and λ_{it}.

Insert Table 3 and Figure 2 about here.

Moving to correlations, Table 4 provides correlations between TFP, product appeal, markups and log prices. Again, we demean TFP and product appeal, as well as log prices, because these measures do not compare much across 8-digit products. The first striking feature emerging from this Table is the strong negative correlation between quantity TFP and product appeal. This is robust to refining the correlation analysis by industry that we accomplish in Figure 3. Indeed, 3 shows a strong (within product) negative correlation between a_{it} and λ_{it} in each of the 9 industries we consider. Furthermore, the very same pattern emerges if we use the DLG-WLD procedure to recover a_{it} and μ_{it} and subsequently use (11) to get an estimate of product appeal (see Figure 4). Interestingly, using very different data and methodology, Jaumandreu and Yin (2017) also come to the conclusion that TFP and demand heterogeneity are negatively correlated.[24]

Insert Table 4 and Figures 3 and 4 about here.

This finding is suggestive of a trade-off between the appeal/quality of a firm's products (as measured by λ) and their production cost (as measured by a) as suggested in, among others, Ackerberg et al. (2007). 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 (higher λ) and so the equilibrium price of Mercedes will be higher than the Nissan price. To be a bit more 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 face very different demands which leads to different prices as well different markups. Both plants are profitable and perhaps generate a very similar revenue productivity (we will come back to this issue below). Yet, their business model is quite different and we are able to identify these differences within our framework.[25]

The second thing worth noting in Table 4 is that both markups and prices are correlated with quantity TFP and product appeal. Table 5 offers more structured insights into the relationship of markups and prices with heterogeneity in demand and TFP. Let's consider markups first. In a world in which the fundamental drivers of heterogeneity across firms are a and λ we would expect markups μ to vary across firms only to the extent that λ, a and predetermined inputs (with the latter contributing to determine short-term marginal costs) vary across firms. In column 1 of Table 5 we regress μ on a, λ and capital k[26] and add

[24] Jaumandreu and Yin (2017) do not have quantity data but impose a number of assumptions under which, what they refer to as cost and product advantages, can be separately identified using revenue and inputs data as well as data on demand shifters. Our approach requires less assumptions and provides what we believe is a more direct and compelling evidence about the negative correlation between TFP and demand heterogeneity.

[25] The presence of a negative relationship between λ and a can be rationalized in several ways. For example, it could be the outcome of firms optimally differentiating themselves in the quality-cost space and/or what is left after selection has taken place and only firms with high enough a and/or high enough λ survive.

[26] Results are robust to including the labour input. By using one predetermined input only (capital) we

a full battery of product and time dummies. In doing so we find reasonable coefficients in that more productive firms and/or firms producing more appealing products charge higher markups.[27] This is confirmed by running the same regression across industries in Table F-2 in Appendix F. Yet, the R^2 is about 0.6 meaning that there is a considerable amount of unexplained heterogeneity in markups. In this respect our framework does allow, as shown by the generalized CES example provided in Section 2.3, for markups that are not fully determined by a, λ and predetermined inputs and these findings suggest markups are not simply a residual dimension of heterogeneity in the data. At the same time, we acknowledge this might well reflect model mispecifications and/or measurement error but we are not able with available data for this paper to draw the line between these two alternative explanations. We are currently pursuing this line of research with more suitable data for France.

Insert Table 5 about here.

Moving from markups to prices reveals similar yet much more powerful patterns. We find in column 2 of Table 5 that more productive firm charge lower prices while firms selling more appealing products charge higher prices. This is again in line with expectations but the striking feature here is the combined predictive power of a, λ and capital k on log prices. The R^2 is basically 1 and it is obtained by analyzing the variation within products, i.e., in this specification log prices, TFP, product appeal and capital are demeaned by 8-digit Prodcom codes and we just consider year dummies. This is confirmed by running the same regression across industries in Table F-3 in Appendix F.

In order to get further insights we consider in columns 3 and 4 of Table 5 a similar regressions but on time-differenced variables. Coefficients in these two columns are thus identified by within-firm changes in prices, markups, TFP, product appeal and capital. Quite surprisingly, we find coefficients very similar to those of columns 1 and 2. At face value, the TFP coefficient thus means that a 10% increase in productivity translates into a 0.037 higher markup and a 8.87% price reduction. The latter implies a cost pass-trough elasticity on prices of 0.887 which is in the ballpark of figures provided in Campa and Goldberg (2005). Looking at industry-specific results in Table F-3 provides a range for the pass-trough elasticity in between 0.8 and 1.

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 6

better capture the overall role of firm size.

[27]The fact that more productive firms charge a higher markup does depend on the features of the underlying demand. In this respect our finding is in line with preferences featuring increasing relative love for variety (from which pro-competitive effects come from) and the presence of market distortions such that the market leads to too little selection with respect to the social optimum. See Dhingra and Morrow (2012) and Zhelobodko et al. (2012) for further details.

reports estimations by industry. In each case we regress a and λ, as well as μ, on their respective time lag. As far as a and λ are concerned, they are both characterized by a high degree of time persistency with autoregressive coefficients being around 0.75-1.00 and an R^2 in the ballpark of 0.8 or above. This evidence supports our choice of a simple Markov process for a and, most importantly, λ. Also note that the autoregressive nature of the productivity and product appeal processes survives the inclusion of firm fixed effects, as indicated by results reported in Table F-4 in Appendix F using the within estimator. In this respect we restate what we believe is an important point. λ captures consumers' overall evaluation of a firm's products quality and appeal; something that arguably does not change dramatically from one year to another. It takes years of effort and costly investments to firms to establish their brand and build their customers' base very much like it takes years of effort and costly investments to firms to put in place and develop an efficient production process for their products. We thus believe there are profound similarities between the processes of productivity and product appeal and so if the former can be approximated by a Markov process we do not see why the latter could not be. Incidentally, the autoregressive coefficients of a and λ are often very close to each other meaning that the MUOMEGA approach to recover a composite measure of quantity TFP and product appeal does find some support in the data. Last, we also look for the sake of completeness at the degree of time persistence of μ which is actually very similar in strength and precision to what we have seen for a and λ. Yet, we have not imposed a Markov process for μ because the assumption of cost-minimization is enough to identify markups.

Insert Table 6 about here.

6.2 Comparison to other methodologies

As already indicated above the correlation of our quantity TFP measure with the one obtained with the DLG-WLD procedure is extremely high both across all observations as well as within products. Turning to markups, we use the same technique adopted in De Loecker et al. (2016) and based on Hall (1986) to retrieve them and so, besides negligible differences, they are also very strongly correlated within the two approaches. De Loecker et al. (2016) does allow for the presence of demand heterogeneity but it is not helpful in quantifying this source of heterogeneity. Our framework allows measuring demand heterogeneity and in this respect it is similar in spirit to Foster et al. (2008).

In their seminal paper, Foster et al. (2008) use production data of US manufacturing firms, containing information on both value and physical quantity, to estimate quantity-based TFP as well as demand heterogeneity. 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 is uncorrelated with demand heterogeneity. We instead assume that demand heterogeneity follow a Markov process while not imposing restrictions on the correlation with TFP.

In light of our framework, the 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 $-\eta_{it} = -\frac{\mu_{it}}{\mu_{it}-1}$ where η_{it} is the (perceived) elasticity of demand.

2. Product appeal and TFP are strongly correlated with each other: 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. Equation (11) provides us with sufficient means to measure demand heterogeneity once we have 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.[28] Figure 5 shows a plot of λ and FHS demand residuals for each industry. The two sets of demand heterogeneity measures are positively and significantly correlated (correlation equal to 0.097) only in industry 1. For the rest, they are completely orthogonal to each other.

Insert Figure 5 about here.

To be fair, our λ 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 a model where log quantity is regressed over log price within our framework. For example, from (12) we have $q_{it} = -\eta_{it}p_{it} + (\eta_{it} - 1)\lambda_{it} - \eta_{it} \ln \kappa_t$ and the residual component is thus $(\eta_{it} - 1)\lambda_{it} - \eta_{it} \ln \kappa_t$ rather than simply λ_{it}. We do not observe $\ln \kappa_t$ but we do observe $q_{it} + \eta_{it}p_{it} = (\eta_{it} - 1)\lambda_{it} - \eta_{it} \ln \kappa_t$. Figure 6 shows a plot of the "modified λ" (as measured by $q_{it} + \eta_{it}p_{it}$) and FHS demand residuals. The situation is now improved with 5 industries in which the two measures are positively and significantly correlated to each other (correlation in between 0.15-0.30). Yet, they overall hardly look like the same measure.

[28]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 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 as well as 8-digit product 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 nine industries.

Insert Figure 6 about here.

7 Decomposing revenue productivity

In what follows we make use of the revenue productivity decomposition provided in (15) and (16) to analyze standard measures of revenue productivity used in the literature (OLS, Olley and Pakes as well as De Loecker and Warzynski). We then use the decomposition to get deeper and sharper insights into productivity questions. More specifically, we decompose the impact of Chinese imports competition on firm revenue productivity into changes in quantity TFP, product appeal, markups and production scale. Finally, we explore the revenue productivity advantage of importing firms, documented in many studies, and assess where it comes from.

7.1 What does revenue productivity measure?

The TFP concept refers to a setting where quantity is used as a measure of output. Yet, quite often quantity and price data is not available to researchers and revenue, or value-added, is instead used as a measure of output. Besides the bias affecting production function parameters in this situation, i.e., the above discussed output price bias (Klette and Griliches, 1996), a more fundamental question is perhaps what, if anything, do revenue-based measures of productivity actually measure. Within our setting revenue productivity is a meaningful measure in that it summarizes the different elements of firm heterogeneity (quantity TFP, product appeal, markups and scale) into a unique indicator as provided in (15).

At face value, correctly measuring revenue productivity $TFP^R_{it} \equiv r_{it} - \bar{q}_{it}$ requires quantity data because this data is needed in order to properly estimate production function coefficients and so scale $\bar{q}_{it}$. A possible way out from this heavy data requirement is to use the MUOMEGA model provided in Section 3.3 and related additional assumptions. A rather different approach would instead consist in just using a standard measure of revenue productivity, and so do not care much about the output price bias, while bearing in mind that revenue productivity would in any case be a synthetic measure conflating many different heterogeneities. As a matter of fact, in our data we actually find some support for the latter approach. More specifically, the revenue productivity measure obtained using our production function coefficients (TFP^R_{it}) is strongly correlated (correlation of 0.85 and above) with other measures of revenue productivity: OLS revenue TFP as well as revenue TFP computed using the methodologies developed in Olley and Pakes (1996) and De Loecker and Warzynski (2012). Furthermore, we show in Table 7 that these measures of revenue productivity correlate well, and in a meaningful way, with the underlying heterogeneities across firms. More specifically, we regress OLS revenue TFP, Olley and Pakes revenue TFP (OP) and De Loecker and Warzynski revenue TFP (DLW) on markups-adjusted TFP ($\tilde{a}$), product appeal ($\tilde{\lambda}$) and scale ($\tilde{\tilde{q}}$) by industry. If we were to run the same regression using TFP_{it}^{R} as a y variable we would get an R^2 of one and 3 coefficients equal to 1 for $\tilde{a}$, $\tilde{\lambda}$ and $\tilde{\tilde{q}}$. In using more common measures of revenue productivity as a y variables we still find large R^2s and overall positive and significant coefficients for quantity TFP, product appeal and scale.

Insert Table 7 about here.

7.2 An application to Chinese imports competition

The decomposition sheds new light on a well studied question. More specifically, numerous studies have explored the many, besides the well-documented negative effects on employment (Autor et al., 2013), impacts of the spectacular rise of Chinese trade, started well before China joined the WTO in 2001, on both developed and developing countries firms and workers. One particular aspect we are interested in is how China has affected the productivity of European firms and in particular Belgian firms. In this respect, Bloom et al. (2016) provide evidence supporting the claim that import competition from China caused an increase in technical change, as well as an increase in revenue TFP, for European firms selling products most affected by rising imports from China. Bloom et al. (2016) rationalize these effects via a number of channels relating competition to innovation.

Bloom et al. (2016) deal with the presence of unobserved demand and/or supply shocks potentially correlated with Chinese imports competition patterns by focusing on a specific industry ("Textile and Apparel") and exploiting detailed information on import quotas. These quotas were imposed at the EU-level on Chinese imports, as well as on imports from other non-WTO countries, and affected some 6-digit products within the industry but not others. As a consequence of China joining the WTO, these quotas were removed over the time frame of our analysis. To provide some context, when these quotas were abolished this generated a 240% increase in Chinese imports on average within the affected product groups. The underlying identifying assumption of this strategy is that unobserved demand/technology shocks are uncorrelated with the strength of quotas to non-WTO countries (like China) in 2000. Since these quotas were built up from the 1950s, and their phased abolition negotiated in the late 1980s was in preparation for the Uruguay Round, Bloom et al. (2016) conclude that this seems like a plausible assumption.

We first start by replicating some key findings of Bloom et al. (2016) and other papers; namely that employment decreased and revenue productivity increased for firms more affected by import competition. We thus match our product-level quota measure $QUOTA_{CPA6}$ to firms in the "Textile and Apparel" industry and run a regression where the time change in either log firm-level labour expenditure or log revenue productivity is used as outcome variable. Results are reported in columns 1 and 2 of Table 8. The negative (positive) and significant coefficient for labour (revenue TFP) indicates that, on average, labour expenditure growth (revenue productivity growth) has been 3.6% lower (0.7% higher) per year over our time frame for firms affected by the quota removal as compared to non-affected firms. Within our framework we can ask a deeper question and in particular how the increase in revenue productivity has materialised. We do so by using markups-adjusted TFP ($\tilde{a}$), product appeal ($\tilde{\lambda}$) and scale ($\tilde{\bar{q}}$) as additional y variables in columns 3 to 5 of Table 8. In this respect note that, by construction, the sum of the 3 coefficients equals the coefficient of revenue TFP (0.0074).

Insert Table 8 about here.

The overall picture emerging from looking at coefficients is suggestive of the following scenario. First, in the light of our model a quota removal is a negative demand shock that should impact product appeal. Indeed, the coefficient of $\tilde{\lambda}$ is significantly negative and quite large. At the same time, firms have reacted to this negative shock by investing massively in cost-reducing innovation and increasing quantity TFP $\tilde{a}$ as indicated by the related positive and large coefficient. Incidentally, the two opposing effects roughly cancel each other out and so the observed increased in revenue TFP essentially comes from the reduction in firm operations/scale $\bar{q}$, i.e., from the increase in the markups-adjusted scale $\tilde{\bar{q}}_{it} = \frac{(1-\mu_{it})\bar{q}_{it}}{\mu_{it}}$.

Quite surprisingly, we reach the very same conclusions using a completely different regression design in Table 9. More specifically, building on Autor et al. (2013) we consider all industries and construct a time-varying 6-digit product-specific measure of Chinese imports penetration in the EU15 market ($IPC^{EU15}_{CPA6,t}$) based on import shares. In order to deal with the presence of unobserved demand/technology shocks at the product-time level characterizing the EU15 market, and correlated with $IPC^{EU15}_{CPA6,t}$, we then instrument this measure with the equivalent Chinese imports penetration measure in the US market $IPC^{US}_{CPA6,t}$. In doing so we also allow for firm fixed effects and year dummies.

Insert Table 9 about here.

Coming back to Table 9, where we provide results based on non-markups-adjusted measures as well as firm markups to complete the picture, one can appreciate in columns 3 and 4 the same two countervailing effects of Chinese imports competition on quantity TFP (positive) and products appeal (negative). At the same time production scale is negatively and significantly affected (column 5) while markups do not display any significant pattern (column 6). Overall, this adds up again to an increase in revenue productivity stemming from higher import penetration (column 2) while the impact on firm employment is negative as in previous studies (column 1). Last but not least, test statistics suggest $IPC^{US}_{CPA6,t}$ is a strong instrument for $IPC^{EU15}_{CPA6,t}$.

7.3 An application to the productivity advantage of importers

As a second application of our decomposition approach, we analyse the links between importing and productivity. Firms operating internationally are typically found to be more productive than purely domestic firms. The empirical evidence on this is vast and covers both exporting and importing firms as well as other forms of international activity (see Bernard et al., 2012a). With particular reference to importing firms, the standard approach in the literature to rationalize the simultaneous presence of importing and non-importing firms, as well as the observed (revenue) productivity advantage of the former, is a combination of heterogeneity in productivity and fixed costs. In particular, only the most productive firms find it profitable to pay sunk import costs. This basic mechanism has been subsequently developed in a number of ways including the choice of sourcing within or outside firm boundaries (Antras and Helpman, 2004) as well as the presence of complementarities across source markets (Antràs et al., 2017).

Crucially, while these models feature heterogeneity in quantity TFP all of the evidence we have so far is in terms of revenue TFP. We first start by showing that the usual positive correlation between revenue-based TFP and firm import status holds in our data. Table 10 reports results of a linear probability model where the dependent variable is a dummy indicating firm import status while revenue-based DLW TFP is the x variable along with a full set of year and product dummies. The last column of Table 10 provides estimates across all industries while in the rest of the Table we show estimates by industry. As one can notice, in all but one case the coefficient of DLW TFP is positive and significant.

Insert Table 10 about here.

Yet, within our framework we can go further and ask whether and how this result materializes. More specifically, we run a similar set of regressions in Table 11 but now using quantity TFP, product appeal and scale as regressors. In doing so we find the following. First, the strongest predictor of import status is scale, i.e., the size of the firm. Second, as far as quantity TFP and product appeal are concerned they equally contribute to further draw the boundary between importing and non-importing firms. In particular, a firm is more likely to be an importer if its quantity TFP is high and/or if its product appeal is high. This suggests that demand heterogeneity is at least as important as differences in underlying physical productivity to understand firm participation into international sourcing. In this respect, it is important to note that our goal here is neither to draw any causal relationships nor to develop a fully fledged model of import participation but rather to uncover correlations that might be useful to further theoretical contributions.

Insert Table 11 about here.

8 Conclusions

We provide a novel framework that simultaneously allows recovering heterogeneity in productivity, demand, and markups across firms while leaving the correlation among the three unrestricted. We accomplish this by explicitly introducing demand heterogeneity and by systematically exploiting assumptions of previous firm-level productivity estimation approaches. We apply our econometric framework to Belgian manufacturing firms and quantify productivity, markups and demand heterogeneity. We show how these heterogeneities 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 gaining deeper and sharper insights on two key questions: firm response to increasing import competition from China and the productivity advantage of importers.

Our methodology is rich enough to be applied to markets where products have some features of both horizontal and vertical differentiation. It allows for multi-product firms, alternative hypotheses on preferences and market structure as well as on the production function and processes for productivity and demand. At the same time, our framework is parsimonious enough to allow retrieving productivity, demand, and markups heterogeneity with relatively little information compared to demand systems models. It also builds upon firm-level data on physical production that is becoming increasingly available to researchers (Belgium, Brazil, Chile, Denmark, France, India, UK and the US to name a few countries). Both elements provide a wide scope of applications to our framework. Furthermore, our approach does not simply allow recovering a consistent measure of TFP while having some other heterogeneities in the background. It also enables measuring all of the three heterogeneities we consider and potentially confront them with many research questions. At the same time, our exact decomposition of revenue productivity in terms of the underlying heterogeneities bridges the gap between quantity and revenue productivity estimations.

Our analysis has policy implications both at the micro and macro level. At the micro level it makes a big 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 our framework allows analyzing aggregate revenue productivity cycles, such as the severe downturn of EU countries' 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 demand. This is the object of ongoing research.

References

Ackerberg, D., Benkard, C. L., Berry, S., and Pakes, A. (2007). Econometric tools for analyzing market outcomes. *Handbook of econometrics*, 6:4171–4276.

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83(6):2411–2451.

Anderson, S. P., De Palma, A., and Thisse, J. F. (1992). *Discrete choice theory of product differentiation*. MIT press.

Antràs, P., Fort, T. C., and Tintelnot, F. (2017). The margins of global sourcing: Theory and evidence from u.s. firms. *American Economic Review*, forthcoming.

Antras, P. and Helpman, E. (2004). Global sourcing. *Journal of Political Economy*, 112(3):552–580.

Atkeson, A. and Burstein, A. (2008). Pricing-to-market, trade costs, and international relative prices. *The American Economic Review*, 98(5):1998–2031.

Autor, D., Dorn, D., and Hanson, G. H. (2013). The china syndrome: Local labor market effects of import competition in the us. *American Economic Review*, 103(6):2121–68.

Behrens, K., Corcos, G., and Mion, G. (2013). Trade crisis? what trade crisis? *Review of economics and statistics*, 95(2):702–709.

Behrens, K., Mion, G., Murata, Y., and Südekum, J. (2014). Trade, wages, and productivity. *International Economic Review*, 55(4):1305–1348.

Berhold, M. H. (1973). The use of distribution functions to represent utility functions. *Management Science*, 19(7):825–829.

Bernard, A., Jensen, B., Redding, S., and Schott, P. (2012a). The empirics of firm heterogeneity and international trade. *Annual Review of Economics*, 4:283–313.

Bernard, A. B., Blanchard, E. J., Van Beveren, I., and Vandenbussche, H. Y. (2012b). Carry-along trade. NBER Working Papers 18246.

Bernard, A. B., Redding, S. J., and Schott, P. K. (2011). Multiproduct firms and trade liberalization. *The Quarterly Journal of Economics*, 126(3):1271–1318.

Berry, S., Levinsohn, J., and Pakes, A. (2004). Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of Political Economy*, 112(1):68–105.

Bloom, N., Draca, M., and Van Reenen, J. (2016). Trade induced technical change? the impact of chinese imports on innovation, it and productivity. *The Review of Economic Studies*, 83(1):87–117.

Caliendo, L., Mion, G., Opromolla, L. D., and Rossi-Hansberg, E. (2015). Productivity and organization in portuguese firms. NBER Working Papers 21811.

Campa, J. M. and Goldberg, L. S. (2005). Exchange rate pass-through into import prices.

The review of Economics and Statistics, 87(4):679–690.

De Loecker, J. (2011). Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity. *Econometrica*, 79(5):1407–1451.

De Loecker, J., Fuss, C., and Van Biesebroeck, J. (2014). International competition and firm performance: Evidence from belgium. NBB Working Papers 269.

De Loecker, J. and Goldberg, P. K. (2014). Firm performance in a global market. *Annu. Rev. Econ.*, 6(1):201–227.

De Loecker, J., Goldberg, P. K., Khandelwal, A. K., and Pavcnik, N. (2016). Prices, markups, and trade reform. *Econometrica*, 84(2):445–510.

De Loecker, J. and Warzynski, F. (2012). Markups and firm-level export status. *American Economic Review*, 102(6):2437–71.

Dhingra, S. and Morrow, J. (2012). Monopolistic competition and optimum product diversity under firm heterogeneity. *Mimeo.*

Di Comite, F., Thisse, J.-F., and Vandenbussche, H. (2014). Verti-zontal differentiation in export markets. *Journal of International Economics*, 93(1):50–66.

Doraszelski, U. and Jaumandreu, J. (2013). R&d and productivity: Estimating endogenous productivity. *The Review of Economic Studies*, 80(4):1338–1383.

Feenstra, R. C. (2003). A homothetic utility function for monopolistic competition models, without constant price elasticity. *Economics Letters*, 78(1):79–86.

Foster, L., Haltiwanger, J., and Syverson, C. (2008). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review*, 98(1):394–425.

Hall, R. E. (1986). Market structure and macroeconomic fluctuations. *Brookings papers on economic activity*, 2:285–338.

Hottman, C. J., Redding, S. J., and Weinstein, D. E. (2016). Quantifying the sources of firm heterogeneity. *The Quarterly Journal of Economics*, 131(3):1291–1364.

Jacob, N. and Mion, G. (2017). The productivity advantage of cities: quantity, quality, markups and scale. University of Sussex Mimeo.

Jaumandreu, J. and Yin, H. (2017). Cost and product advantages: A firm-level model for the chinese exports and industry growth. CEPR Discussion Paper 11862.

Klette, T. J. and Griliches, Z. (1996). The inconsistency of common scale estimators when output prices are unobserved and endogenous. *Journal of Applied Econometrics*, 11(4):343–361.

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

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725.

Mion, G. and Zhu, L. (2013). Import competition from and offshoring to china: A curse or blessing for firms? *Journal of International Economics*, 89(1):202–215.

Muûls, M. (2015). Exporters, importers and credit constraints. *Journal of International Economics*, 95(2):333 – 343.

Muûls, M. and Pisu, M. (2009). Imports and Exports at the Level of the Firm: Evidence from Belgium. *The World Economy*, 32(5):692–734.

Nocke, V. and Schutz, N. (2016). Multiproduct-firm oligopoly: An aggregative games approach. CEPR Discussion Paper 11539.

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):l263–l297.

Roberts, M. J., Xu, D. Y., Fan, X., and Zhang, S. (2016). The role of firm factors in demand, cost, and export market selection for chinese footwear producers. *Mimeo*.

Rodríguez-López, J. A. (2011). Prices and exchange rates: A theory of disconnect. *The Review of Economic Studies*, 78(3):1135–1177.

Spence, M. (1976). Product selection, fixed costs, and monopolistic competition. *The Review of Economic Studies*, 43(2):217–235.

Wooldridge, J. M. (2009). On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters*, 104(3):112–114.

Zhelobodko, E., Kokovin, S., Parenti, M., and Thisse, J.-F. (2012). Monopolistic competition: Beyond the constant elasticity of substitution. *Econometrica*, 80(6):2765–2784.

Table 1: Basic summary statistics of the estimation sample

Industry	Industry description	Nace	N Obs	Statistic	Revenue	Quantity	Labour	Materials	Capital
1	Food prod, beverages and tobacco	15+16	1,317						
				mean	1.431	14.311	-0.220	0.979	-0.024
				st. dev.	1.165	1.741	1.003	1.240	1.336
				p5	-0.277	11.786	-1.626	-0.944	-2.376
				p95	3.574	17.458	1.689	3.205	2.209
2	Textiles and leather	17 to 19	1,225						
				mean	1.047	12.256	-0.359	0.578	-1.035
				st. dev.	1.129	1.938	0.995	1.227	1.549
				p5	-0.696	9.029	-1.789	-1.284	-3.650
				p95	3.147	15.677	1.589	2.793	1.535
3	Wood except furniture	20	348						
				mean	1.284	11.359	-0.166	0.812	-0.394
				st. dev.	1.354	2.827	1.169	1.457	1.598
				p5	-0.314	7.738	-1.505	-0.951	-2.962
				p95	4.579	15.778	2.586	4.325	2.721
4	Pulp, paper, publish. and print.	21+22	975						
				mean	1.916	14.875	0.396	1.447	-0.299
				st. dev.	1.240	1.679	1.190	1.296	1.719
				p5	0.124	12.087	-1.181	-0.530	-3.353
				p95	4.227	17.728	2.735	3.829	2.610
5	Chemicals and rubber	24+25	1,043						
				mean	1.865	14.060	0.145	1.453	0.206
				st. dev.	1.121	2.378	1.085	1.162	1.267
				p5	0.272	10.062	-1.336	-0.308	-1.870
				p95	3.852	17.819	2.249	3.481	2.273
6	Other non-metallic mineral prod.	26	1,215						
				mean	1.792	16.002	0.295	1.298	0.239
				st. dev.	1.039	2.850	1.111	1.049	1.312
				p5	0.310	10.795	-1.258	-0.225	-2.021
				p95	3.774	19.318	2.447	3.121	2.491
7	Metals and fabric. metal prod.	27+28	2,814						
				mean	1.114	12.563	-0.160	0.588	-0.827
				st. dev.	0.980	2.301	0.875	1.067	1.262
				p5	-0.162	8.151	-1.246	-0.909	-2.883
				p95	3.036	16.109	1.543	2.668	1.360
8	Machin., electr. and optic. equip.	29 to 33	1,108						
				mean	1.514	8.828	0.233	1.002	-0.693
				st. dev.	1.169	3.365	1.031	1.259	1.439
				p5	-0.070	3.989	-1.078	-0.759	-3.049
				p95	3.506	14.374	2.105	3.127	1.738
9	Transport equipment and n.e.c.	34 to 36	1,055						
				mean	1.244	9.552	-0.110	0.762	-0.830
				st. dev.	1.130	2.935	1.072	1.206	1.419
				p5	-0.276	5.361	-1.489	-0.951	-3.411
				p95	3.481	16.102	1.981	3.158	1.371

Notes: Revenue denotes log revenue, quantity denotes log quantity in the unit specific to a product, labour denotes log of labour expenditure, materials denotes log of materials expenditure, capital denotes log capital stock. All monetary values are expressed in current million euros.

Table 2: Estimates of production function parameters with our procedure

Industry	Description	Labour	Materials	Capital	γ
1	Food products, beverages and tobacco	0.397^{a}	0.728^{a}	0.045^{a}	1.169^{a}
		(0.029)	(0.040)	(0.014)	(0.061)
2	Textiles and leather	0.325^{a}	0.636^{a}	0.020^{c}	0.981^{a}
		(0.020)	(0.019)	(0.012)	(0.014)
3	Wood except furniture	0.340^{a}	0.632^{a}	0.026	0.998^{a}
		(0.050)	(0.049)	(0.021)	(0.058)
4	Pulp, paper, publishing and printing	0.427^{a}	0.629^{a}	0.070^{a}	0.986^{a}
		(0.065)	(0.092)	(0.017)	(0.141)
5	Chemicals and rubber	0.328^{a}	0.648^{a}	0.034^{c}	1.010^{a}
		(0.040)	(0.052)	(0.019)	(0.071)
6	Other non-metallic mineral products	0.316^{a}	0.622^{a}	0.047^{a}	0.985^{a}
		(0.039)	(0.051)	(0.015)	(0.078)
7	Basic metals and fabric. metal prod.	0.338^{a}	0.629^{a}	0.024^{a}	0.991^{a}
		(0.015)	(0.012)	(0.008)	(0.005)
8	Machinery, electric. and optical equip.	0.347^{a}	0.630^{a}	0.026^{b}	1.004^{a}
		(0.033)	(0.023)	(0.011)	(0.008)
9	Transport equipment and n.e.c.	0.313^{a}	0.636^{a}	0.025	0.974^{a}
		(0.032)	(0.031)	(0.016)	(0.039)

Notes: γ denotes returns to scale. Bootstrapped standard errors in parenthesis (200 replications). [a] p<0.01, [b] p<0.05, [c] p<0.1.

Table 3: Standard deviation of TFP, product appeal and markups by industry

Industry	Description	TFP	product appeal	markups
1	Food products, beverages and tobacco	0.416	0.477	0.154
2	Textiles and leather	0.604	0.671	0.130
3	Wood except furniture	0.843	0.914	0.180
4	Pulp, paper, publishing and printing	0.775	0.843	0.152
5	Chemicals and rubber	0.952	0.970	0.079
6	Other non-metallic mineral products	0.520	0.607	0.123
7	Basic metals and fabric. metal prod.	0.860	0.896	0.169
8	Machinery, electric. and optical equip.	0.917	0.925	0.139
9	Transport equipment and n.e.c.	1.021	1.020	0.151

Notes: TFP and product appeal are demeaned by 8-digit Prodcom codes.

Table 4: Correlations between TFP, product appeal (λ), markups and log prices

	TFP	λ	markups	prices
TFP	1			
λ	-0.9680^a	1		
markups	-0.0789^a	0.1730^a	1	
prices	-0.9940^a	0.9670^a	0.0807^a	1

Notes: TFP, product appeal and prices are demeaned by 8-digit Prodcom codes. ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1.

Table 5: Regression of markups and (demeaned) log prices on TFP, product appeal (λ) and log capital

Estimation method	OLS		First differences	
Dependent Variable	Markups	Prices	Markups	Prices
TFP	0.3424^a	-0.9093^a	0.3724^a	-0.8866^a
	(0.0193)	(0.0080)	(0.0246)	(0.009)
λ	0.3570^a	0.0735^a	0.3707^a	0.1048^a
	(0.0185)	(0.0076)	(0.0246)	(0.0089)
capital	-0.0252^a	-0.0101^a	0.0037	-0.0046^c
	(0.0030)	(0.0012)	(0.0029)	(0.0019)
Year dummies	Yes	Yes	Yes	Yes
Prod dummies	Yes	No	Yes	No
N Obs	11,100	11,100	7,768	7,768
R^2	0.6338	0.9878	0.3971	0.9925

Notes: Prices, TFP, product appeal and capital are demeaned by 8-digit Prodcom codes in the second and fourth regressions. The first two columns correspond to OLS estimations while columns 3 and 4 correspond to OLS on first-differenced variables. Bootstrapped standard errors in parenthesis (200 replications). ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1.

Table 6: OLS regression of TFP, product appeal (λ) and markups on their time lag by industry

Industry	1	2	3	4	5	6	7	8	9
					TFP				
lag TFP	0.9743^a	0.9718^a	0.9865^a	0.9577^a	0.8715^a	0.9711^a	0.8665^a	0.7482^a	0.8332^a
	(0.0155)	(0.0126)	(0.0157)	(0.0231)	(0.0264)	(0.0114)	(0.0245)	(0.0579)	(0.0355)
N Obs	901	843	232	710	702	867	2,000	785	738
R^2	0.8742	0.8785	0.962	0.8986	0.7867	0.9371	0.7035	0.5986	0.73
					λ				
lag λ	0.9654^a	0.9688^a	0.9886^a	0.9532^a	0.8737^a	0.9503^a	0.8769^a	0.7465^a	0.8292^a
	(0.0136)	(0.0136)	(0.012)	(0.0323)	(0.0265)	(0.0153)	(0.0247)	(0.059)	(0.0413)
N Obs	901	843	232	710	702	867	2,000	785	738
R^2	0.8763	0.8825	0.9688	0.871	0.7813	0.8942	0.7216	0.6001	0.7239
					markup				
lag markup	0.9476^a	0.9464^a	0.9242^a	0.9504^a	0.9327^a	0.9344^a	0.9006^a	0.8717^a	0.9742^a
	(0.0139)	(0.0162)	(0.0372)	(0.0178)	(0.0261)	(0.0193)	(0.0143)	(0.02)	(0.0336)
N Obs	901	843	232	710	702	867	2,000	785	738
R^2	0.9112	0.8724	0.8637	0.8965	0.8762	0.8878	0.8018	0.7837	0.8133

Notes: Time dummies are included in estimations but are not reported here. Bootstrapped standard errors in parenthesis (200 replications). ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1.

Table 7: Regression of OLS revenue TFP, Olley and Pakes revenue TFP and De Loecker and Warzynski revenue TFP on markups-adjusted TFP ($\tilde{a}$), product appeal ($\tilde{\lambda}$) and scale ($\tilde{\tilde{q}}$) by industry

Industry	1	2	3	4	5	6	7	8	9
					OLS revenue TFP				
$\tilde{a}$	0.1529^{a}	0.6517^{a}	0.7786^{a}	0.3242^{a}	0.2783^{a}	0.6023^{a}	0.8696^{a}	0.8806^{a}	0.7055^{a}
	(0.0134)	(0.0164)	(0.0222)	(0.0218)	(0.0135)	(0.0128)	(0.0057)	(0.0178)	(0.0158)
$\tilde{\lambda}$	0.1666^{a}	0.6639^{a}	0.7869^{a}	0.3428^{a}	0.2950^{a}	0.6201^{a}	0.8776^{a}	0.8878^{a}	0.7223^{a}
	(0.0136)	(0.0166)	(0.0223)	(0.0222)	(0.0139)	(0.0132)	(0.0057)	(0.0179)	(0.0163)
$\tilde{\tilde{q}}$	0.0761^{a}	0.7170^{a}	0.8476^{a}	0.3166^{a}	0.3308^{a}	0.6055^{a}	0.8894^{a}	0.8814^{a}	0.7586^{a}
	(0.0125)	(0.0148)	(0.0223)	(0.0248)	(0.0191)	(0.0159)	(0.0050)	(0.0167)	(0.0187)
N Obs	1,317	1,225	348	975	1,055	1,215	2,814	1,108	1,055
R^2	0.7611	0.9567	0.9688	0.8122	0.9233	0.9515	0.9766	0.9743	0.8955
					Olley and Pakes revenue TFP				
$\tilde{a}$	0.1389^{a}	0.6420^{a}	0.7311^{a}	0.3210^{a}	0.2606^{a}	0.6214^{a}	0.8667^{a}	0.8821^{a}	0.7114^{a}
	(0.0119)	(0.0178)	(0.0312)	(0.0233)	(0.0166)	(0.0120)	(0.0068)	(0.0184)	(0.0155)
$\tilde{\lambda}$	0.1546^{a}	0.6545^{a}	0.7361^{a}	0.3392^{a}	0.2773^{a}	0.6384^{a}	0.8742^{a}	0.8893^{a}	0.7278^{a}
	(0.0120)	(0.0180)	(0.0313)	(0.0237)	(0.0170)	(0.0124)	(0.0069)	(0.0186)	(0.0161)
$\tilde{\tilde{q}}$	0.0708^{a}	0.7099^{a}	0.8299^{a}	0.3134^{a}	0.3271^{a}	0.6259^{a}	0.8857^{a}	0.8843^{a}	0.7626^{a}
	(0.0115)	(0.0165)	(0.0306)	(0.0259)	(0.0207)	(0.0162)	(0.0062)	(0.0167)	(0.0185)
N Obs	1,317	1,225	348	975	1,055	1,215	2,814	1,108	1,055
R^2	0.7522	0.9536	0.9408	0.8058	0.9165	0.9549	0.9691	0.9743	0.8993
					De Loecker and Warzynski revenue TFP				
$\tilde{a}$	0.0246	0.6946^{a}	0.0128	0.4004^{a}	0.2730^{a}	0.6934^{a}	0.9099^{a}	0.8530^{a}	0.7230^{a}
	(0.0201)	(0.0278)	(0.3747)	(0.0263)	(0.0137)	(0.0122)	(0.0053)	(0.0448)	(0.0141)
$\tilde{\lambda}$	0.0269	0.6945^{a}	-0.0931	0.4084^{a}	0.2926^{a}	0.7048^{a}	0.9148^{a}	0.8522^{a}	0.7388^{a}
	(0.0204)	(0.0282)	(0.3730)	(0.0268)	(0.0142)	(0.0124)	(0.0054)	(0.0448)	(0.0146)
$\tilde{\tilde{q}}$	-0.0969^{a}	0.7490^{a}	0.7327^{b}	0.3928^{a}	0.2359^{a}	0.6652^{a}	0.9236^{a}	0.8604^{a}	0.7704^{a}
	(0.0165)	(0.0280)	(0.3474)	(0.0377)	(0.0197)	(0.0157)	(0.0048)	(0.0422)	(0.0176)
N Obs	1,317	1,225	348	975	1,055	1,215	2,814	1,108	1,055
R^2	0.7071	0.9168	0.7932	0.6465	0.9201	0.9091	0.9836	0.9222	0.9052

Notes: Time and 8-digit product dummies are included in estimations but are not reported here. Bootstrapped standard errors in parenthesis (200 replications). ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1.

Table 8: Disentangling the impact of Chinese imports competition on revenue productivity in terms of markups-adjusted TFP ($\tilde{a}$), product appeal ($\tilde{\lambda}$) and scale ($\tilde{\bar{q}}$): quota analysis on the "Textile and Apparel" industry

Outcome measure	Labour	Rev.TFP	$\tilde{a}$	$\tilde{\lambda}$	$\tilde{\bar{q}}$
$Quota_{CPA6}$	-0.0360^{c}	0.0074^{b}	0.1095^{c}	-0.1163^{b}	0.0143^{b}
	(0.0207)	(0.0033)	(0.0593)	(0.0571)	(0.0064)
Observations	700	700	700	700	700
R-squared	0.0052	0.0055	0.0033	0.0039	0.0079

Notes: $Quota_{CPA6}$ denotes the share of products within a CPA6 code belonging to the "Textile and Apparel" industry affected by a removal of quota on Chinese imports. Firm-level clustered standard errors in parenthesis. ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1.

Table 9: Disentangling the impact of Chinese imports competition on revenue productivity in terms of TFP (a), product appeal (λ), scale ($\bar{q}$) and markups (μ): Chinese import penetration analysis

Outcome measure	Labour	Rev.TFP	a	λ	$\bar{q}$	μ
$IPC^{EU15}_{CPA6,t}$	-0.7393^{a}	0.1847^{a}	1.3205^{a}	-1.0728^{a}	-0.8483^{a}	0.0493
	(0.2412)	(0.0423)	(0.4028)	(0.4140)	(0.2574)	(0.0650)
Observations	10,161	10,161	10,161	10,161	10,161	10,161
R-squared	0.1741	0.0292	0.0174	0.0111	0.2141	0.0369
Firm FE and year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Kleibergen-Paap rk LM statistic under-id	146.71	146.71	146.71	146.71	146.71	146.71
P-value	0	0	0	0	0	0
Kleibergen-Paap rk Wald F statistic weak id.	226.34	226.34	226.34	226.34	226.34	226.34

Notes: Instrumental variable estimator with firm fixed effects implemented. Chinese import penetration in the EU15 market ($IPC^{EU15}_{CPA6,t}$) is instrumented with Chinese import penetration in the US market ($IPC^{US}_{CPA6,t}$). Firm-level clustered standard errors in parenthesis. ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1. The Kleibergen-Paap rk LM statistic tests for under-identification. P-value reported. The Kleibergen-Paap rk Wald F statistic tests for weak identification. The corresponding Stock-Yogo weak ID test critical value is 16.38 for a 10% maximal IV size bias.

Table 10: The revenue productivity advantage of importers by industry: De Loecker and Warzynski (DLW) revenue TFP

Industry	1	2	3	4	5	6	7	8	9	All
DLW revenue TFP	2.3207^{a}	1.1475^{a}	0.5283^{a}	0.9602^{a}	0.1967	2.3030^{a}	0.6613^{a}	1.2348^{a}	0.4739^{b}	0.7752^{a}
	(0.1541)	(0.2095)	(0.0381)	(0.1632)	(0.2972)	(0.3459)	(0.1394)	(0.2419)	(0.2244)	(0.0450)
Observations	1,317	1,225	348	975	1,055	1,215	2,814	1,108	1,055	11,112
R-squared	0.4685	0.3137	0.5749	0.6471	0.3465	0.2996	0.3212	0.4636	0.3015	0.4245

Notes: The dependent variable is a dummy indicating whether a firm is an importer at time t. A linear probability model is estimated. Time and 8-digit product dummies are included in estimations but are not reported here. Bootstrapped standard errors in parenthesis (200 replications). ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1.

Table 11: Disentangling the revenue productivity advantage of importers by industry in terms of TFP (a), product appeal (λ) and scale ($\bar{q}$)

Industry	1	2	3	4	5	6	7	8	9	All
a	0.0870^b	0.0914	-0.0601	0.0324	0.6182^a	0.2420^a	0.0094	0.1802^a	0.3554^a	0.1069^a
	(0.0413)	(0.0781)	(0.1645)	(0.0379)	(0.1057)	(0.0384)	(0.0335)	(0.0581)	(0.0654)	(0.0165)
λ	-0.0429	0.0876	-0.0652	0.0581^c	0.6141^a	0.1554^a	-0.0063	0.1493^b	0.3309^a	0.0877^a
	(0.0354)	(0.0706)	(0.1564)	(0.0315)	(0.1029)	(0.0275)	(0.0321)	(0.0580)	(0.0643)	(0.0154)
$\bar{q}$	0.2421^a	0.1779^a	0.2286^a	0.1067^a	0.1870^a	0.2578^a	0.2970^a	0.2468^a	0.2616^a	0.2238^a
	(0.0083)	(0.0161)	(0.0247)	(0.0104)	(0.0202)	(0.0104)	(0.0081)	(0.0185)	(0.0114)	(0.0040)
N Obs	1,317	1,225	348	975	1,055	1,215	2,814	1,108	1,055	11,112
R^2	0.6090	0.3886	0.5895	0.6728	0.4298	0.4400	0.5042	0.5514	0.5324	0.5419

Notes: The dependent variable is a dummy indicating whether a firm is an importer at time t. A linear probability model is estimated. Time and 8-digit product dummies are included in estimations but are not reported here. Bootstrapped standard errors in parenthesis (200 replications). ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1.

Figure 1: The importance of heterogeneity in demand across firms

Figure 2: Distribution of markups by industry

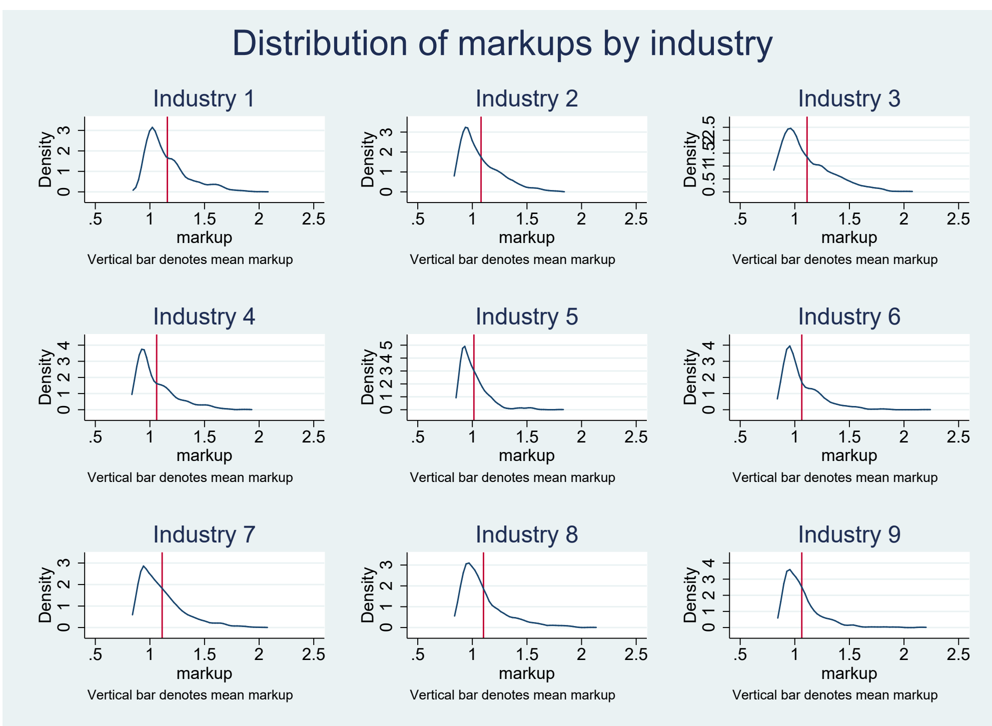


Figure 3: Within 8-digit products correlation between TFP and product appeal by industry: our estimation procedure

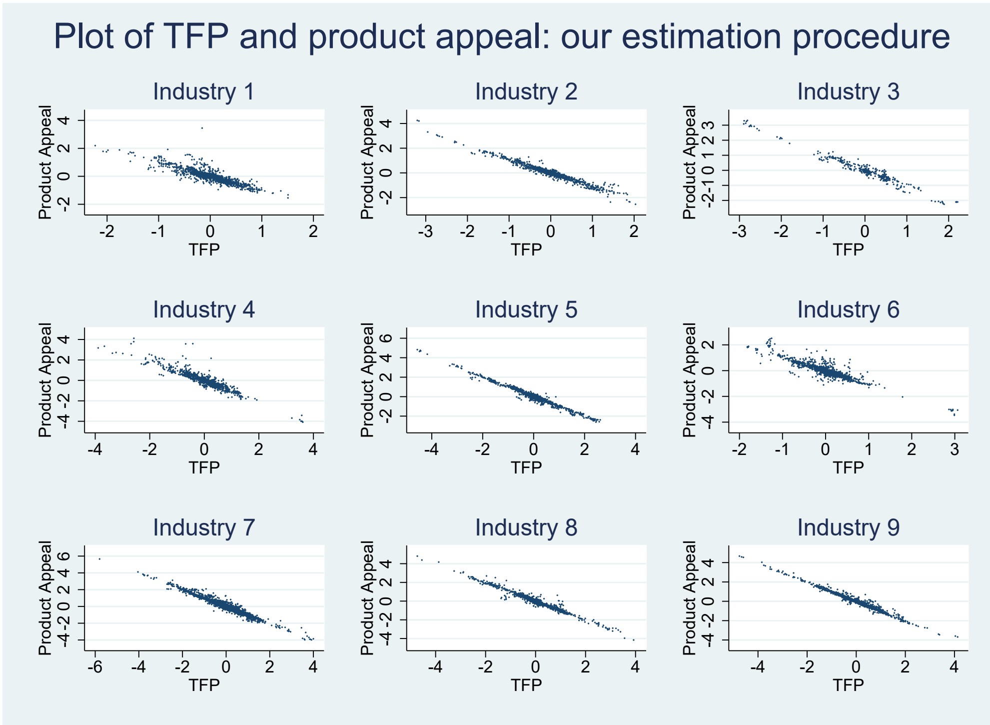


Figure 4: Within 8-digit products correlation between TFP and product appeal by industry: DLG-WLD procedure

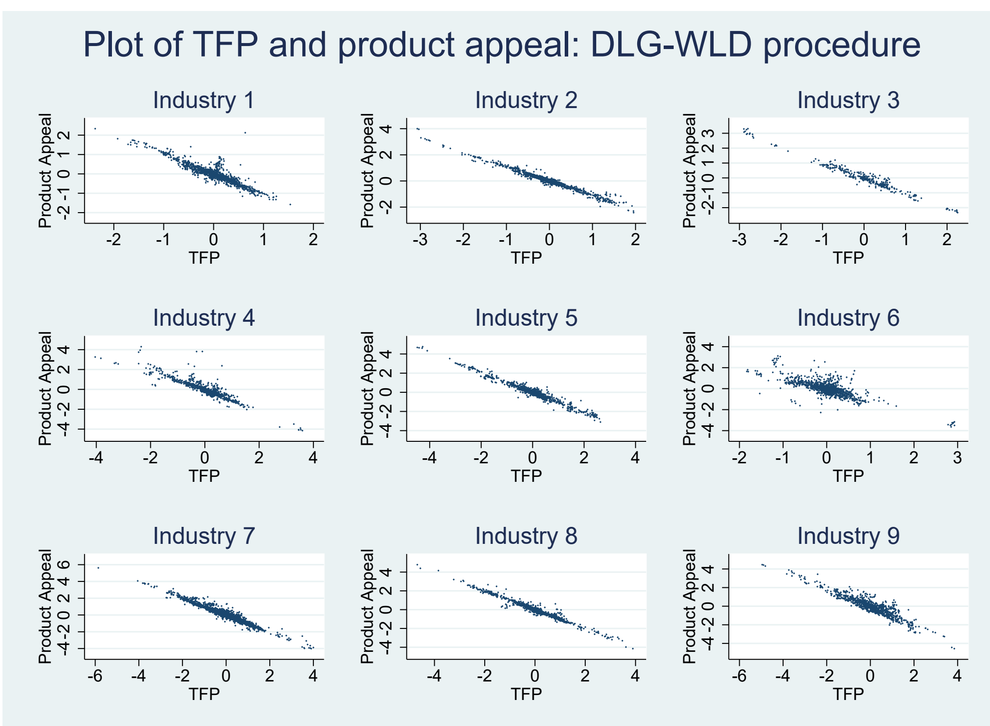


Figure 5: Plot of λ and FHS demand residual

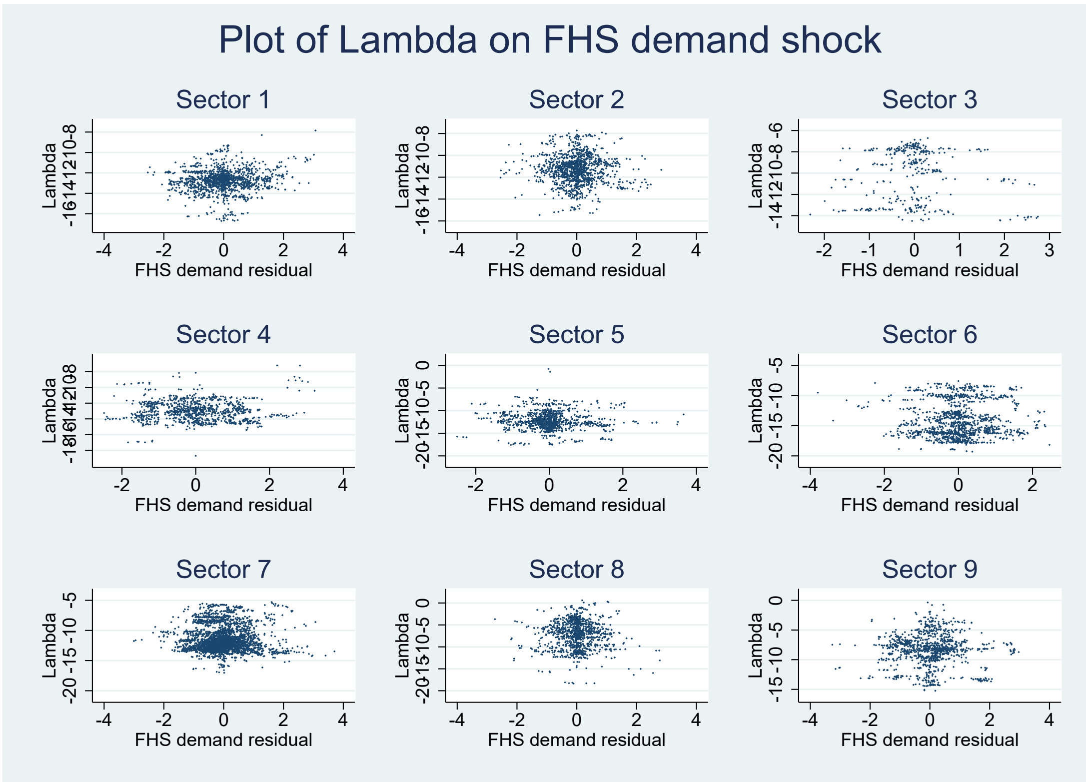
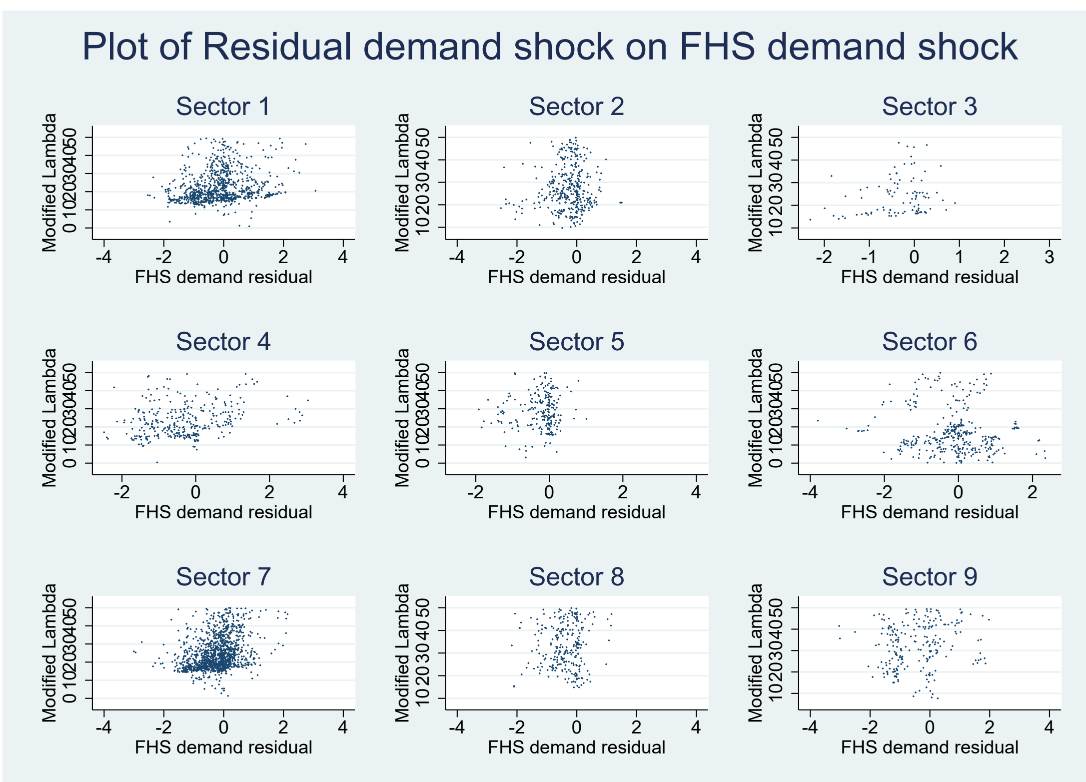


Figure 6: Plot of modified λ and FHS demand residual



Appendix

In Appendices A to D we show how to interpret and apply the key property (9) to various demand and market structure settings as well as how to extend our analysis to a wider set of production functions, to richer processes for productivity and product appeal and to multi-product firms. In Appendix E we derive the conditional input demand for materials and show how invertibility does not generally hold.

A The key property (9)

The key property we need is that the elasticity of revenue with respect to quantity is proportional to the elasticity of revenue with respect to product appeal. In what follows we show how to interpret and apply the key property (9) to various demand and market structure settings. In Section A.1 we consider monopolistic competition models with a representative consumer while in Section A.2 we move to discrete/continuous choice models. We extend our analysis to oligopoly in Section A.3 where we build upon Atkeson and Burstein (2008) while in Section A.4 we propose a departure from the first-order linear approximation assumption and derive a modified procedure for a non log-linear demand.

A.1 Monopolistic competition models with a representative consumer

(9) is equivalent to requiring that the elasticity of prices with respect to quantity differs from the elasticity of prices with respect to product appeal by one: $\frac{\partial p_{it}}{\partial \lambda_{it}} = \frac{\partial p_{it}}{\partial q_{it}} + 1$. This is straightforward to obtain and interpret in monopolistic competition models with a representative consumer. Consider a representative consumer who maximises at each point in time t a differentiable utility function $U(.)$ subject to budget B_t:

$$\max_{Q} \left\{ U\left(\tilde{Q}\right)\right\} \text{ s.t. } \int_i P_{it} Q_{it} \mathrm{d}i - B_t = 0 \tag{A-1}$$

where $\tilde{Q}$ is a vector of elements $\Lambda_{it} Q_{it}$. Therefore, while the representative consumer chooses quantities Q, these quantities enter into the utility function as $\tilde{Q}$ and Λ_{it} can be interpreted as a measure of quality of a particular variety. 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 $\Lambda_{it} = \overline{\Lambda}$ or $\overline{\Lambda}$ more units of a a variety with $\Lambda_{it} = 1$. The first order conditions of the utility maximization problem imply:

$$\frac{\partial U}{\partial Q_{it}} = \frac{\partial U}{\partial \tilde{Q}_{it}} \frac{\partial \tilde{Q}_{it}}{\partial Q_{it}} = \frac{\partial U}{\partial \tilde{Q}_{it}} \Lambda_{it} = \kappa_t P_{it},$$

where κ_t is a Lagrange multiplier and $\frac{\partial \tilde{Q}_{it}}{\partial Q_{it}} = \Lambda_{it}$. Taking logs we have:

$$\ln \frac{\partial U}{\partial \tilde{Q}_{it}} + \lambda_{it} = \ln \kappa_t + p_{it}. \tag{A-2}$$

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 $U(.)$, this might be tricky to work out. Nonetheless, equation (A-2) already tells us a lot about the shape of such demand functions. On the one hand, differentiating both sides with respect to q_{it} yields:

$$\frac{\partial p_{it}}{\partial q_{it}} = \frac{\partial \ln \frac{\partial U}{\partial \tilde{Q}_{it}}}{\partial q_{it}} = \frac{\partial \ln \frac{\partial U}{\partial \tilde{Q}_{it}}}{\partial \tilde{q}_{it}} \frac{\partial \tilde{q}_{it}}{\partial q_{it}} = \frac{\partial \ln \frac{\partial U}{\partial \tilde{Q}_{it}}}{\partial \tilde{q}_{it}}, \tag{A-3}$$

where $\frac{\partial \tilde{q}_{it}}{\partial q_{it}} = 1$ and $\frac{\partial p_{it}}{\partial q_{it}} \equiv -\frac{1}{\eta_{it}}$. On the other hand, keeping in mind that $\frac{\partial \tilde{q}_{it}}{\partial \lambda_{it}} = 1$, differentiation of both sides with respect to λ_{it} gives:

$$\frac{\partial p_{it}}{\partial \lambda_{it}} = \frac{\partial \ln \frac{\partial U}{\partial \tilde{Q}_{it}}}{\partial \lambda_{it}} + 1 = \frac{\partial \ln \frac{\partial U}{\partial \tilde{Q}_{it}}}{\partial \tilde{q}_{it}} \frac{\partial \tilde{q}_{it}}{\partial \lambda_{it}} + 1 = \frac{\partial \ln \frac{\partial U}{\partial \tilde{Q}_{it}}}{\partial \tilde{q}_{it}} + 1 = 1 - \frac{1}{\eta_{it}},$$

i.e., the elasticity of the price with respect to quantity differs from the elasticity of the price with respect to product appeal by one. Using equation (A-2) we can write log revenue r_{it} (up to a constant) as:

$$r_{it} = p_{it} + q_{it} = \ln \frac{\partial U}{\partial \tilde{Q}_{it}} + \lambda_{it} + q_{it} = \ln \frac{\partial U}{\partial \tilde{Q}_{it}} + \tilde{q}_{it}.$$

Differentiating both sides with respect to $\tilde{q}_{it}$ and making use of equation (A-3) we have:

$$\frac{\partial r_{it}}{\partial \tilde{q}_{it}} = \frac{\partial \ln \frac{\partial U}{\partial \tilde{Q}_{it}}}{\partial \tilde{q}_{it}} + 1 = -\frac{1}{\eta_{it}} + 1 = \frac{1}{\mu_{it}},$$

and so we finally get:

$$\Delta r_{it} \approx \frac{1}{\mu_{it}} \Delta \tilde{q}_{it} = \frac{1}{\mu_{it}} \Delta(q_{it} + \lambda_{it}). \tag{A-4}$$

Therefore, for any preferences structure that can be used to model monopolistic competition and that can be described by a well-behaved differentiable direct utility we can, starting from the baseline formulation $U(Q)$, introduce quality in such a way that equation (A-4) is satisfied. The advantage of equation (A-4) is that it can be directly used for estimations without the need to explicitly solve for the demand functions of the different varieties.

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

$$U(\tilde{Q}) = \int_{i\in I_t} a_{it}\left(\tilde{Q}_{it}\right)^{b_{it}} \mathrm{d}i = \int_{i\in I_t} a_{it}\Lambda_{it}^{b_{it}}\left(Q_{it}\right)^{b_{it}} \mathrm{d}i,$$

where $b_{it} = 1 - \frac{1}{\eta_{it}}$. If we further impose $a_{it} = \frac{\eta_{it}}{\eta_{it}-1}$ not only equation (A-4) holds but we actually get equation (13). Other examples of preferences falling within our class include the CARA preferences used in Behrens et al. (2014) as well as the Translog preferences featuring in Feenstra (2003) and Rodríguez-López (2011). Contrary to CARA preferences in the case of the Translog there is no closed-form expression for the utility function and the demand system is derived directly from the expenditure function. Yet the utility function does exists and it is well behaved and so it falls within our class. A workable Translog framework can be readily obtained by considering that everything works as if firms sell quantities $\tilde{Q}_{it} = Q_{it}\Lambda_{it}$ while charging prices $\tilde{P}_{it} = P_{it}/\Lambda_{it}$. Also note that $\tilde{P}_{it}\tilde{Q}_{it} = P_{it}Q_{it}$ by construction and so total consumer expenditure B_t in terms of $\tilde{Q}_{it}$ is the same as in terms of Q_{it}. Therefore, one simply needs to substitute log prices $\tilde{p}_{it}$ with $p_{it} - \lambda_{it}$ in the Translog expenditure function.

A.2 Discrete/continuous choice models

Discrete/continuous choice models represent a generalisation of standard discrete choice models including, for example, the Multinomial Logit. They are obtained from a random utility framework in which consumers not only choose one alternative amongst many but also how much to consume of a particular good. In what follows we borrow the terminology from Nocke and Schutz (2016) to which the reader is referred for more details.

Let I_t be the set of differentiated products. A discrete/continuous choice model is a collection of functions of individual prices $\{H_{it}(P_{it})\}_{i\in I_t}$, where, for every i in I_t

- H_{it} is C^3 from $\mathbb{R}_{++}$ to $\mathbb{R}_{++}$
- $V_{it}' < 0$, $V_{it}'' \geq 0$,

where $V_{it} \equiv \log(H_{it})$ is an indirect sub-utility function in a world in which only product i and the outside good, good 0, are available.

The consumer makes at each point in time t choices as follows. He first observes idiosyncratic random components $\{\varepsilon_{it}\}_{i\in I_t}$ that are iid type-1 extreme-value as well as prices $\{P_{it}\}_{i\in I_t}$. He then chooses only one product and, if he chooses product i, he consumes $D_{it}(P_{it}) = -V_{it}'(P_{it})$ units of that product (Roy's identity) and uses the rest of his income to consume the outside good. In doing so he receives indirect utility $Y + V_{it}(P_{it}) + \varepsilon_{it}$. The consumer chooses the product i that maximizes indirect utility, i.e., $i \in \arg\max_{j\in I_t}\{Y + V_{jt}(P_{jt}) + \varepsilon_{jt}\}$.

By Holman and Marley's theorem the probability of choosing alternative i is $\mathbb{P}_{it}(P) = \frac{e^{V_{it}(P_{it})}}{\sum_{j \in I_t} e^{V_{jt}(P_{jt})}}$ where P is the vector of prices while the expected demand for product i is given by:

$$Q_{it}(P) = \frac{-H_{it}^{'}(P_{it})}{\sum_{j \in I_t} H_{jt}(P_{jt})}. \tag{A-5}$$

The demand system (A-5) satisfies the IIA property and the basic Multinomial Logit is obtained as a special case by setting $H_{it}(P_{it}) = H(P_{it}) = e^{-P_{it}}$. The basic Multinomial Logit can be enriched by introducing a measure of quality ($H_{it}(P_{it}) = e^{\Lambda_{it} - P_{it}}$) but it would not in general satisfy (9). This is due to the fact that the consumer only chooses one unit of a given product and so our concept of quality does not fit within this framework. (A-5) allows to go beyond the Multionomail Logit and, by relaxing the assumption of unit consumption, (A-5) can be used to characterize preferences satisfying (9). For example, an equivalent to generalized CES preferences can be obtained from (A-5) by setting $H_{it}(P_{it}) = \Lambda_{it}^{\eta_{it}-1} P_{it}^{1-\eta_{it}}$. In this case we have:

$$Q_{it}(P) = \frac{(\eta_{it} - 1)\Lambda_{it}^{\eta_{it}-1} P_{it}^{-\eta_{it}}}{\sum_{j \in I_t} \Lambda_{jt}^{\eta_{jt}-1} P_{jt}^{1-\eta_{jt}}}.$$

In the limit case of monopolistic competition (the set I_t is large enough) the denominator is fixed for an individual firm and so it can be readily verified that (9) holds. More broadly (A-5) can be used to generate demand systems compatible with (9) by an appropriate choice of $H_{it}(P_{it})$. This comes from the equivalence in terms of aggregate demand between representative consumer's models and discrete choice models laid down in Anderson et al. (1992). With specific reference to oligopolistic competition, Nocke and Schutz (2016) build on (A-5) to develop an oligopoly Bertrand competition model featuring product heterogeneity in quality Λ_{it} that can be casted within our framework. The game they consider is aggregative and they are able to establish conditions for both existence and uniqueness of a price equilibrium.

A.3 Oligopoly: Atkeson and Burstein (2008)

We provide here an example of how (9) holds within an oligopoly model based on Atkeson and Burstein (2008) and further refined in Hottman et al. (2016) for multi-product firms. The key ingredient is the same as in the monopolistic competition case, namely that quantities enter into preferences as $\tilde{Q}_{it} = \Lambda_{it} Q_{it}$. Atkeson and Burstein (2008) consider a nested CES model of quantity competition à la Cournot in which firms sell differentiated varieties and are large enough to perceive their impact on industry aggregates while charging markups that depend upon their market share. More specifically, a finite number of single-product firms operates within each industry j where preferences are characterized by a CES demand with parameter η_j. At each point in time final consumption is produced by a competitive firm using the output of a continuum of industries $\mathbb{Q}_{jt}$ for $j \in [0,1]$ as inputs subject to a CES production function with parameter η and $1 < \eta < \eta_j$, i.e., varieties within an industry are more substitutable with each other than industry outputs Q_{jt} across industries. Contrary to the monopolistic competition case, firm i operating in industry j does recognize that sectoral prices and quantities vary when it changes its quantity. Introducing product appeal within this framework is quite straightforward.

First, industry j aggregate output at time t is:

$$\mathbb{Q}_{jt} = \left(\sum_{i \in I_{jt}} (Q_{ijt}\Lambda_{ijt})^{\frac{\eta_j - 1}{\eta_j}} \right)^{\frac{\eta_j}{\eta_j - 1}}, \tag{A-6}$$

where I_{jt} is the set of varieties (firms) available within industry j at time t, Q_{ijt} is firm i output in industry j and Λ_{ijt} is product appeal. The inverse demand corresponding to varieties within an industry is:

$$\frac{P_{ijt}}{\mathbb{P}_{jt}} = \left(\frac{Q_{ijt}}{\mathbb{Q}_{jt}} \right)^{-\frac{1}{\eta_j}} \Lambda_{ijt}^{\frac{\eta_j - 1}{\eta_j}}, \tag{A-7}$$

where $\mathbb{P}_{jt}$ is the CES price index for industry j and and P_{ijt} is firm i price in industry j. Firm revenue is thus $R_{ijt} = P_{ijt}Q_{ijt}$. The inverse demand corresponding to industry outputs is instead:

$$\frac{\mathbb{P}_{jt}}{\mathbb{P}_t} = \left(\frac{\mathbb{Q}_{jt}}{\mathbb{Q}_t} \right)^{-\frac{1}{\eta}}, \tag{A-8}$$

where $\mathbb{P}_t$ and $\mathbb{Q}_t$ are the CES price and quantity indexes for the whole economy and in particular the latter is:

$$\mathbb{Q}_t = \left(\sum_j (\mathbb{Q}_{jt})^{\frac{\eta - 1}{\eta}} \right)^{\frac{\eta}{\eta - 1}}.$$

In choosing the optimal quantity firms realize the impact of their choices on industry aggregates and in particular on $\mathbb{Q}_{jt}$. They have instead a zero measure at the aggregate level and so take $\mathbb{P}_t$ and $\mathbb{Q}_t$ as given. Combining (A-7) and (A-8) the relevant demand is:

$$\frac{P_{ijt}}{\mathbb{P}_t} = \left(\frac{Q_{ijt}}{\mathbb{Q}_{jt}} \right)^{-\frac{1}{\eta_j}} \Lambda_{ijt}^{\frac{\eta_j - 1}{\eta_j}} \left(\frac{\mathbb{Q}_{jt}}{\mathbb{Q}_t} \right)^{-\frac{1}{\eta}}. \tag{A-9}$$

Using the properties of CES demand, the market share of firm i in industry j ($s_{Rijt} \equiv R_{ijt}/(\mathbb{P}_{jt}\mathbb{Q}_{jt})$) equals the elasticity of the industry quantity index with respect to firm quantity:

$$\frac{\partial \mathbb{Q}_{jt}}{\partial Q_{ijt}} \frac{Q_{ijt}}{\mathbb{Q}_{jt}} = s_{Rijt}. \tag{A-10}$$

Symmetrically we have:

$$\frac{\partial \mathbb{Q}_{jt}}{\partial \Lambda_{ijt}} \frac{\Lambda_{ijt}}{\mathbb{Q}_{jt}} = s_{Rijt}. \tag{A-11}$$

From (A-9) and (A-10), the elasticity of firm price P_{ijt} with respect to firm quantity Q_{ijt} is thus $-\frac{1}{\eta_j}(1 - s_{Rijt}) - \frac{1}{\eta} s_{Rijt}$. Therefore, the elasticity of demand is:

$$\epsilon_{it} = -\frac{\partial q_{ijt}}{\partial p_{ijt}} = \left(\frac{1}{\eta_j}(1 - s_{Rijt}) + \frac{1}{\eta} s_{Rijt} \right)^{-1}, \tag{A-12}$$

while from profit maximization the markup μ_{ijt} is related to the elasticity of demand in the usual way:

$$\mu_{ijt} = \frac{\epsilon_{ijt}}{\epsilon_{ijt} - 1}. \tag{A-13}$$

Multiplying both sides of (A-9) by quantity delivers the revenue equation:

$$\frac{R_{ijt}}{\mathbb{P}_t} = (Q_{ijt}\Lambda_{ijt})^{\frac{\eta_j - 1}{\eta_j}} \mathbb{Q}_{jt}^{\frac{1}{\eta_j}} \left(\frac{\mathbb{Q}_{jt}}{\mathbb{Q}_t} \right)^{-\frac{1}{\eta}}, \tag{A-14}$$

from which the elasticity of revenue with respect to quantity is equal to the elasticity of revenue with respect to product appeal and equal to one over the profit maximizing markup:

$$\frac{\partial r_{ijt}}{\partial q_{ijt}} = \frac{\partial r_{ijt}}{\partial \lambda_{ijt}} = \frac{\eta_j - 1}{\eta_j} + \frac{1}{\eta_j} s_{Rijt} - \frac{1}{\eta} s_{Rijt} = \frac{1}{\mu_{ijt}}. \tag{A-15}$$

As shown in Hottman et al. (2016), the above framework including product appeal can be generalized to the case of multi-product firms while allowing for a rich structure featuring an endogenous number of products for each firm and cannibalization effects while remaining quite analytically tractable. The key tool to make this happen, and in particular to avoid complications arising from cannibalization effects, is to add yet another CES nest at the level of the firm. More specifically, the different products produced by a multi-product firm within an industry/product group are bundled together into a firm-level CES aggregate. These firm-level CES aggregates are then bundled again into a CES product group-level aggregate with an elasticity of substitution that is smaller than the one corresponding to firm-level CES aggregates, i.e., products produced by a firm are more substitutable with each other than products sold by different firms.

A.4 Modified procedure for a non log-linear demand

Here we discuss how our model can be extended beyond (9), i.e., without resorting to any linear approximation by fully specifying a preference structure and working out the corresponding algebra for the revenue equation. We are particularly interested in a flexible structure leading to a non-log linear demand. Specifically, we look here at an additively separable utility function shaped like the Gaussian CDF:[29]

$$U(\tilde{q}) = \int_{i \in I_t} \Phi\left(\tilde{q}_{it}, \beta_0, \beta_1, \beta_2\right) \mathrm{d}i,$$

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

$$\Phi\left(\tilde{q}_{it}\right) = u\left(\tilde{q}_{it}\right) = \int_{-\infty}^{\tilde{q}_{it}} \phi\left(\tilde{q}_{it}\right) \mathrm{d}\tau.$$

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

$$\frac{\phi\left(\tilde{q}_{it}\right)}{Q_{it}\kappa_t} = P_{it},$$

where $\phi\left(\tilde{q}_{it}\right)$ is the Gaussian PDF. In what follows we focus on monopolistic competition and set $\phi\left(\tilde{q}_{it}\right) = \exp\left(-\beta_2^2 \tilde{q}_{it}^3 + \beta_1 \tilde{q}_{it} + \beta_0\right)$ 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 ignore terms that are constant across firms and so have:

$$r_{it} = \ln \phi\left(\tilde{q}_{it}\right),$$

as well as:

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

as expected from (7).

Using the definitions $\chi_{it} \equiv s_{Mit}\left(m_{it} - k_{it}\right)$ and $\chi'_{it} \equiv \alpha_L\left(l_{it} - k_{it}\right) + \gamma k_{it}$ we obtain:

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

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

$$= \left(\frac{1}{3} + \frac{2}{3}\beta_1 \mu_{it}\right)\chi_{it} + \left(\frac{1}{3}\frac{1}{\mu_{it}\beta_1} + \frac{2}{3}\right)\beta_1 \chi'_{it} + \left(\frac{1}{3}\frac{1}{\mu_{it}\beta_1} + \frac{2}{3}\right)\beta_1 (a_{it} + \lambda_{it})$$

[29]See Berhold (1973) for further discussion of the Gaussian CDF as a utility function.

$$\Rightarrow \frac{r_{it}}{\frac{1}{3}\frac{1}{\mu_{it}\beta_1}+\frac{2}{3}} = \frac{\frac{1}{3}+\frac{2}{3}\beta_1\mu_{it}}{\frac{1}{3}\frac{1}{\mu_{it}\beta_1}+\frac{2}{3}}\chi_{it} + \beta_1\chi'_{it} + \beta_1(a_{it}+\lambda_{it}).$$

Further note that: $\frac{\frac{1}{3}+\frac{2}{3}\beta_1\mu_{it}}{\frac{1}{3}\frac{1}{\mu_{it}\beta_1}+\frac{2}{3}} = \frac{1+2\beta_1\mu_{it}}{\frac{1}{\mu_{it}\beta_1}+2} = \beta_1\mu_{it} = \beta_1\frac{\alpha_M}{s_{Mit}}$. Hence we can write:

$$\frac{r_{it}}{\frac{1}{3}\frac{s_{Mit}}{\beta_1\alpha_M}+\frac{2}{3}} = \beta_1\alpha_M\frac{\chi_{it}}{s_{Mit}} + \beta_1\chi'_{it} + \beta_1(a_{it}+\lambda_{it}),$$

which implies:

$$r_{it} = \left(\frac{1}{3}\frac{1}{\mu_{it}} + \frac{2}{3}\beta_1\right)(q_{it}+\lambda_{it}). \tag{A-16}$$

Equation (A-16) is the equivalent of equation (13). Building on the same logic utilized for equations (19) and (20) one finally gets:

$$\begin{aligned}
\widetilde{LHS}_{it} &= \beta_1\alpha_L(l_{it}-k_{it}) + \beta_1\gamma k_{it} + \phi_a\widetilde{LHS}_{it-1} - \phi_a\beta_1\alpha_L(l_{it-1}-k_{it-1}) - \phi_a\beta_1\gamma k_{it-1} \\
&+ \frac{(\phi_\lambda-\phi_a)}{\beta_1}\left(\frac{r_{it-1}}{\frac{1}{3}\frac{s_{Mit-1}}{\beta_1\alpha_M}+\frac{2}{3}}\right) - (\phi_\lambda-\phi_a)\beta_1 q_{it-1} + \beta_1(\nu_{ait}+\nu_{\lambda it}).
\end{aligned} \tag{A-17}$$

where $\widetilde{LHS}_{it} = \frac{r_{it}}{\frac{1}{3}\frac{s_{Mit}}{\beta_1\alpha_M}+\frac{2}{3}} - \beta_1\alpha_M\frac{\chi_{it}}{s_{Mit}}$. Estimation of the various parameters in equation (A-17) can be carried by non-linear GMM, as in De Loecker and Warzynski (2012) for example, by considering that the error term $u_{it} = \beta_1(\nu_{ait}+\nu_{\lambda it})$ is a function of some data as well as of the parameters and by building 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 our baseline procedure one can avoid exploiting parameters' constraints and extract some reduced-form parameters: $\beta_1\alpha_M$, $\beta_1\alpha_L$ and $\beta_1\gamma$ as well as ϕ_a. In the very same way we recover γ in the baseline procedure via a second step estimation based on the quantity equation, we can, by using estimates of $\beta_1\alpha_M$, $\beta_1\alpha_L$, $\beta_1\gamma$ and ϕ_a, write the quantity equation as a linear expression involving only one unknown parameter (β_1) and one right-hand side variable. Therefore, we can use a simple IV strategy based on the same moment conditions above to identify β_1, and so α_M, α_L and γ, and ultimately productivity, product appeal and markups. As far as product appeal is concerned, (A-16) needs to be used instead of (10) to retrieve them.

Finally notice that

$$\frac{1}{\mu_{it}} = -3\beta_2^2\tilde{q}_{it}^2 + \beta_1.$$

Hence

$$\frac{\partial\left(-3\beta_2^2\tilde{q}_{it}^2+\beta_1\right)}{\partial q_{it}} = -6\beta_2^2\tilde{q}_{it},$$

i.e., the markup depends (positively) on equilibrium quantity.

B More general production functions

Here we show how to 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 γ represents the returns to scale). We also have that

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

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

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

and

$$\begin{aligned} q_{it} &= \frac{\partial q_{it}}{\partial m_{it}} m_{it} + \frac{\partial q_{it}}{\partial l_{it}} l_{it} + \left(\gamma - \frac{\partial q_{it}}{\partial m_{it}} - \frac{\partial q_{it}}{\partial l_{it}}\right) k_{it} \\ &\quad - \frac{1}{2}\alpha_{MM} m_{it}^2 - \frac{1}{2}\alpha_{LL} l_{it}^2 - \frac{1}{2}\alpha_{KK} k_{it}^2 - \alpha_{MK} k_{it} m_{it} - \alpha_{ML} m_{it} l_{it} - \alpha_{LK} l_{it} k_{it} + a_{it}. \end{aligned} \tag{B-1}$$

From the first order conditions

$$s_{Mit}\mu_{it} = \frac{\partial q_{it}}{\partial m_{it}} \tag{B-2}$$

holds so that

$$\begin{aligned} q_{it} &= s_{Mit}\mu_{it} m_{it} + \frac{\partial q_{it}}{\partial l_{it}} l_{it} + \left(\gamma - s_{Mit}\mu_{it} - \frac{\partial q_{it}}{\partial l_{it}}\right) k_{it} \\ &\quad - \frac{1}{2}\alpha_{MM} m_{it}^2 - \frac{1}{2}\alpha_{LL} l_{it}^2 - \frac{1}{2}\alpha_{KK} k_{it}^2 - \alpha_{MK} k_{it} m_{it} - \alpha_{ML} m_{it} l_{it} - \alpha_{LK} l_{it} k_{it} + a_{it}. \end{aligned}$$

Now add λ_{it} to and subtract $\mu_{it} s_{Mit} (m_{it} - k_{it})$ from both sides and note that, because of (10) and (B-2), the first equality also holds:

$$\begin{aligned}
LHS_{it}\frac{\partial q_{it}}{\partial m_{it}} &= (q_{it}+\lambda_{it})-\mu_{it}s_{Mit}(m_{it}-k_{it}) \qquad \text{(B-3)}\\
&= \gamma k_{it}+\frac{\partial q_{it}}{\partial l_{it}}(l_{it}-k_{it})-\frac{1}{2}\alpha_{MM}m_{it}^2-\frac{1}{2}\alpha_{LL}l_{it}^2-\frac{1}{2}\alpha_{KK}k_{it}^2-\alpha_{MK}k_{it}m_{it}\\
&\quad -\alpha_{ML}m_{it}l_{it}-\alpha_{LK}l_{it}k_{it}+a_{it}+\lambda_{it},
\end{aligned}$$

where LHS_{it} is the same as in equation (18), 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 equations (6), (14), (19) and (20) into equation (B-3) (while replacing the old α_M with $\partial q_{it}/\partial m_{it}$) and dividing both sides by γ we get:

$$\begin{aligned}
LHS_{it}\frac{\partial q_{it}}{\partial m_{it}}\frac{1}{\gamma} &= e_{it}-\phi_a e_{it-1}+\partial q_{it}/\partial m_{it}\frac{1}{\gamma}\phi_a LHS_{it-1} \qquad \text{(B-4)}\\
&+ \partial q_{it}/\partial m_{it}\frac{1}{\gamma}(\phi_\lambda-\phi_a)\left(\frac{r_{it-1}}{s_{Mit-1}}-\frac{1}{\partial q_{it}/\partial m_{it}}q_{it-1}\right)+u_{it},
\end{aligned}$$

where $e_{it}=k_{it}-\frac{1}{2}\frac{\alpha_{MM}}{\gamma}m_{it}^2+\frac{\alpha_{LL}}{\gamma}\left(\frac{1}{2}l_{it}^2-l_{it}k_{it}\right)-\frac{1}{2}\frac{\alpha_{KK}}{\gamma}k_{it}^2-\frac{\alpha_{MK}}{\gamma}k_{it}m_{it}-\frac{\alpha_{ML}}{\gamma}m_{it}k_{it}-\frac{\alpha_{LK}}{\gamma}k_{it}^2-\frac{\alpha_L}{\gamma}(l_{it}-k_{it})$ and $u_{it}{=}\frac{1}{\gamma}(\nu_{ait}+\nu_{\lambda it})$.

Estimation of the various parameters in equation (B-4) can be carried by non-linear GMM, as in De Loecker and Warzynski (2012) for example, by considering that u_{it} is a function of some data as well as of the parameters, and builds on moment conditions such as $E[k_{it}u_{it}]=E[k_{it-1}u_{it}]=E[m_{it-1}u_{it}]=E[l_{it-1}u_{it}]=0$ as well as $E\left[m_{it-1}^2u_{it}\right]=E[m_{it-1}k_{it-1}u_{it}]=E[m_{it-1}l_{it-2}u_{it}]=E[k_{it-2}l_{it-2}u_{it}]=0$, and so on and so forth. Considering moments up to $t-2$ ($t-1$) there are 30 (13) such moments conditions that can be exploited. As in the Cobb-Douglas case, it is perhaps best not to exploit parameters' constraints (this mean for example estimating $\frac{\alpha_{MM}}{\gamma}$ rather than trying to separately identify α_{MM} and γ from the revenue equation) and extract some reduced form parameters to be used in a second stage regression based on the quantity equation.

All of the parameters of the quantity equation (B-1) have been identified up to the scaling γ in the previous stage and we can write it as:

$$q_{it} = \gamma z_{it}+a_{it}, \qquad \text{(B-5)}$$

where:

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

As in the Cobb-Douglas case we can further substitute for a_{it} using the same logic of equation (20) (while replacing the old α_M with $\partial q_{it}/\partial m_{it}$) to get to the equivalent of (26) and use the usual moment conditions on a simple linear model with a single regressor to identify γ and ultimately productivity, product appeal and markups.

C More general processes for a and λ

C.1 Non-linear Markov processes, time-invariant unobserved heterogeneity and measurement error in capital

Our model can be easily extended to non-linear Markov processes for a and λ as well as to the presence of time-invariant unobserved heterogeneity and measurement error in capital. Consider, for example, the first case and in particular:

$$\begin{aligned} a_{it} &= \phi_{1a}a_{it-1} + \phi_{2a}a_{it-1}^2 + \nu_{ait} \\ \lambda_{it} &= \phi_{1\lambda}\lambda_{it-1} + \phi_{2\lambda}\lambda_{it-1}^2 + \nu_{\lambda it}. \end{aligned} \tag{C-1}$$

By substituting equations (19), (20) and (C-1) into (18) we obtain:

$$\begin{aligned} LHS_{it} &= \frac{\gamma}{\alpha_M}k_{it} + \frac{\alpha_L}{\alpha_M}(l_{it}-k_{it}) + \phi_{1a}LHS_{it-1} - \phi_{1a}\frac{\gamma}{\alpha_M}k_{it-1} - \phi_{1a}\frac{\alpha_L}{\alpha_M}(l_{it-1}-k_{it-1}) \\ &+ (\phi_{1\lambda}-\phi_{1a})\left(\frac{r_{it-1}}{s_{Mit-1}} - \frac{1}{\alpha_M}q_{it-1}\right) \\ &+ (\phi_{2\lambda}-\phi_{2a})\left(\alpha_M(\frac{r_{it-1}}{s_{Mit-1}})^2 + \frac{1}{\alpha_M}(q_{it-1})^2 - 2(\frac{r_{it-1}}{s_{Mit-1}}q_{it-1})\right) \\ &+ \phi_{2a}\left(\alpha_M(LHS_{it-1})^2 + \frac{\gamma^2}{\alpha_M}(k_{it-1})^2 + \frac{\alpha_L^2}{\alpha_M}(l_{it-1}-k_{it-1})^2\right) \\ &+ \phi_{2a}\left(-2\gamma LHS_{it-1}k_{it-1} - 2\alpha_L LHS_{it-1}(l_{it-1}-k_{it-1}) - 2\alpha_M(LHS_{it-1}\frac{r_{it-1}}{s_{Mit-1}})\right) \\ &+ \phi_{2a}\left(2(LHS_{it-1}q_{it-1}) + 2\gamma(k_{it-1}\frac{r_{it-1}}{s_{Mit-1}}) + 2\frac{\gamma\alpha_L}{\alpha_M}(k_{it-1}(l_{it-1}-k_{it-1})) - 2\frac{\gamma}{\alpha_M}(k_{it-1}q_{it-1})\right) \\ &+ \phi_{2a}\left(2\alpha_L((l_{it-1}-k_{it-1})\frac{r_{it-1}}{s_{Mit-1}}) - 2\frac{\alpha_L}{\alpha_M}((l_{it-1}-k_{it-1})q_{it-1})\right) + \frac{1}{\alpha_M}(\nu_{ait}+\nu_{\lambda it}). \end{aligned} \tag{C-2}$$

Equation (C-2) can be used to estimate $\frac{\gamma}{\alpha_M} \equiv b_1$, $\frac{\alpha_L}{\alpha_M} \equiv b_2$, ϕ_{1a} and ϕ_{2a}[30] by a suitable IV linear regression with a change in variables, $(l_{it-2}-k_{it-2})$ used as instrument for $(l_{it}-k_{it})$ and some reduced-form parameters. In turn these estimates could be employed in

[30] As in the baseline case, b_1, b_2 and ϕ_{1a} can be directly obtained from, respectively, the coefficients of k_{it}, $l_{it}-k_{it}$ and LHS_{it-1} with no need to exploit reduced-form parameters constraints. As for ϕ_{2a}, this can be obtained as 1/2 times the coefficient of the interaction between LHS_{it-1} and q_{it-1}.

the corresponding expression of (25):

$$
\begin{aligned}
q_{it} &= \frac{\gamma \hat{b}_2}{\hat{b}_1}\left(l_{it}-k_{it}\right)+\frac{\gamma}{\hat{b}_1}\left(m_{it}-k_{it}\right)+\gamma k_{it} \\
&+ \hat{\phi}_{1a}\frac{\gamma}{\hat{b}_1}LHS_{it-1}-\hat{\phi}_{1a}\gamma k_{it-1}-\hat{\phi}_{1a}\frac{\gamma \hat{b}_2}{\hat{b}_1}(l_{it-1}-k_{it-1})-\hat{\phi}_{1a}\left(r_{it-1}\frac{\gamma}{\hat{b}_1 s_{Mit-1}}-q_{it-1}\right) \qquad \text{(C-3)} \\
&+ \hat{\phi}_{2a}(\frac{\gamma}{\hat{b}_1})^2(LHS_{it-1})^2+\hat{\phi}_{2a}\gamma^2(k_{it-1})^2+\hat{\phi}_{2a}(\frac{\gamma \hat{b}_2}{\hat{b}_1})^2(l_{it-1}-k_{it-1})^2+\hat{\phi}_{2a}(\frac{\gamma}{\hat{b}_1})^2(\frac{r_{it-1}}{s_{Mit-1}})^2+\hat{\phi}_{2a}(q_{it-1})^2 \\
&- 2\hat{\phi}_{2a}\frac{\gamma^2}{\hat{b}_1}LHS_{it-1}k_{it-1}-2\hat{\phi}_{2a}\hat{b}_2(\frac{\gamma}{\hat{b}_1})^2LHS_{it-1}(l_{it-1}-k_{it-1})-2\hat{\phi}_{2a}(\frac{\gamma}{\hat{b}_1})^2LHS_{it-1}\frac{r_{it-1}}{s_{Mit-1}} \\
&+ 2\hat{\phi}_{2a}\frac{\gamma}{\hat{b}_1}LHS_{it-1}q_{it-1}+2\hat{\phi}_{2a}\frac{\gamma^2\hat{b}_2}{\hat{b}_1}k_{it-1}(l_{it-1}-k_{it-1})+2\hat{\phi}_{2a}\frac{\gamma^2}{\hat{b}_1}k_{it-1}\frac{r_{it-1}}{s_{Mit-1}}-2\hat{\phi}_{2a}\gamma k_{it-1}q_{it-1} \\
&+ 2\hat{\phi}_{2a}\hat{b}_2(\frac{\gamma}{\hat{b}_1})^2(l_{it-1}-k_{it-1})\frac{r_{it-1}}{s_{Mit-1}}-2\hat{\phi}_{2a}\frac{\gamma \hat{b}_2}{\hat{b}_1}(l_{it-1}-k_{it-1})q_{it-1}+\nu_{ait},
\end{aligned}
$$

from which the γ parameter can be 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 γ and the other where the only unknown coefficient is γ^2. As in the baseline case, instrumenting is needed and there are many moments that can be used: capital in t as well as lagged capital, materials, labour, revenue and quantity including square terms and interactions.

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

$$
\begin{aligned}
a_{it} &= \phi_a a_{it-1}+u_{ai}+\nu_{ait} \\
\lambda_{it} &= \phi_\lambda \lambda_{it-1}+u_{\lambda i}+\nu_{\lambda it}. \qquad \text{(C-4)}
\end{aligned}
$$

By substituting equations (19), (20) and (C-4) into equation (18) we obtain:

$$
\begin{aligned}
LHS_{it} &= \frac{\gamma}{\alpha_M}k_{it}+\frac{\alpha_L}{\alpha_M}\left(l_{it}-k_{it}\right)+\phi_a LHS_{it-1}-\phi_a\frac{\gamma}{\alpha_M}k_{it-1}-\phi_a\frac{\alpha_L}{\alpha_M}\left(l_{it-1}-k_{it-1}\right) \\
&+ (\phi_\lambda-\phi_a)\left(\frac{r_{it-1}}{s_{Mit-1}}-\frac{1}{\alpha_M}q_{it-1}\right)+\frac{1}{\alpha_M}\left(u_{ai}+u_{\lambda i}\right)+\frac{1}{\alpha_M}\left(\nu_{ait}+\nu_{\lambda it}\right). \qquad \text{(C-5)}
\end{aligned}
$$

Equation (C-5) is almost identical to (21) with the only difference being the composite unobserved heterogeneity term $\frac{1}{\alpha_M}\left(u_{ai}+u_{\lambda i}\right)$. Therefore, it can be transformed into a linear regression model parallel to (22) with the only difference being that, the simultaneous presence of an unobservable time-invariant component correlated with regressors ($\frac{1}{\alpha_M}\left(u_{ai}+u_{\lambda i}\right)$) and the lag of the dependent variable (LHS_{it-1}), calls for the use of, for example, a dynamic panel data estimator rather than linear IV. Similar arguments apply to the quantity equation (26).

Last but not least the presence of standard measurement error in, for example, capital is relatively straightforward to accommodate in our framework. Such measurement error would imply that the error term in equation (21) is correlated with k_{it} and k_{it-1}. Yet, very much like in Wooldridge (2009), the simple solution to this problem is to add appropriate instruments for k_{it} and k_{it-1} like, for example, inputs, revenue and quantity and time $t-2$. At the same time, the moment conditions $E\{\nu_{ait}k_{it}\} = 0$ and $E\{\nu_{ait}k_{it-1}\} = 0$ used to identify γ in equation (25) would be violated by the presence of measurement error in capital. Yet, there are still enough restrictions to identify γ.

C.2 Endogenous processes

Doraszelski and Jaumandreu (2013) introduce costly R&D investments affecting productivity and optimally chosen by firms who wish to maximize the expected net present value of future cash flows. Within our framework their hypothesis translate into the following productivity process:

$$a_{it} = \phi_a \; a_{it-1} + \phi_{RD} \; RD_{it-1} + \nu_{ait}, \tag{C-6}$$

where RD_{it-1} is R&D expenditure at time $t-1$ and ν_{ait} is uncorrelated with both a_{it-1} and RD_{it-1}. The reasoning behind the zero correlation is as follows. Firm anticipates the effect of R&D on productivity in period t when making the decision about investment in knowledge in period $t-1$. When the decision about investment in knowledge is made in period t-1, the firm is only able to anticipate the expected effect of R&D on productivity in period t as given by $\phi_{RD} \; RD_{it-1}$ while its actual effect also depends on the realization of the productivity innovation ν_{ait} that occurs after the investment has been completely carried out. The productivity innovation ν_{ait} represents the uncertainties that are naturally linked to productivity plus the uncertainties inherent in the R&D process such as chance in discovery, degree of applicability, and success in implementation.

It is straightforward to show that, with data on R&D expenditure at hand, we can build (C-6) into MULAMA. Indeed, accounting for (C-6) simply requires adding RD_{it-1} as an additional covariate in (22) and (26). More broadly, we can consider firms putting in place costly investments affecting productivity (process R&D: $RD^{Process}$) as well as quality (product R&D: $RD^{Product}$). We thus have:

$$\begin{aligned} a_{it} &= \phi_a \; a_{it-1} + \phi_{RD}^{Process} \; RD_{it-1}^{Process} + \nu_{ait} \\ \lambda_{it} &= \phi_\lambda \lambda_{it-1} + \phi_{RD}^{Product} \; RD_{it-1}^{Product} + \nu_{\lambda it}, \end{aligned} \tag{C-7}$$

where ν_{ait} and $\nu_{\lambda it}$ are uncorrelated with covariates. Again, putting (C-7) into MULAMA is straightforward: add both R&D covariates in (22) and only $RD^{Process}$ in and (26).

In the case where such data is not available it is still possible to use MULAMA while adding more structure to the problem. The simplest assumption to make is that the return on investment of $RD^{Process}_{it-1}$ depends upon a_{it-1} while the return on investment of $RD^{Product}_{it-1}$ depends upon λ_{it-1}. If there are quadratic $R\&D$ costs common across firms the optimal investments in process and product R&D will be a function of, respectively, a_{it-1} and λ_{it-1} only. By adding a linearity assumption ($RD^{Process}_{it-1} = \phi^{'}_a\ a_{it-1}$ and $RD^{Product}_{it-1} = \phi^{'}_\lambda\ \lambda_{it-1}$) we have:

$$\begin{aligned} a_{it} &= \phi_a\ a_{it-1} + \phi^{'}_a\ a_{it-1} + \nu_{ait} = \phi^*_a a_{it-1} + \nu_{ait} \\ \lambda_{it} &= \phi_\lambda \lambda_{it-1} + \phi^{'}_\lambda \lambda_{it-1} + \nu_{\lambda it} = \phi^*_\lambda \lambda_{it-1} + \nu_{\lambda it}, \end{aligned} \tag{C-8}$$

where $\phi^*_a = \phi_a + \phi^{'}_a$ and $\phi^*_\lambda = \phi_\lambda + \phi^{'}_\lambda$. (C-8) is isomorphic to (6) and (14) so the baseline estimation procedure can be applied even though the interpretation of the autoregressive coefficients is now different.

A more involved scenario in one in which the return and/or cost of $RD^{Process}_{it-1}$ depends upon both a_{it-1} and λ_{it-1} while the return and/or cost of $RD^{Product}_{it-1}$ also depends upon both a_{it-1} and λ_{it-1}. This would for example capture the link between the two R&D expenditures via the opportunity cost of investing in one or the other. In this case we would have:

$$\begin{aligned} a_{it} &= \phi^*_a\ a_{it-1} + \phi^{'}_\lambda\ \lambda_{it-1} + \nu_{ait} \\ \lambda_{it} &= \phi^*_\lambda \lambda_{it-1} + \phi^{'}_a a_{it-1} + \nu_{\lambda it}. \end{aligned} \tag{C-9}$$

It is still possible to implement (C-9) into the MULAMA model. Combining (18), (19), and (20) with (C-9) one gets the following expression for the revenue equation:

$$\begin{aligned} LHS_{it} &= \frac{\gamma}{\alpha_M} k_{it} + \frac{\alpha_L}{\alpha_M}(l_{it} - k_{it}) + \bar{\phi}_a LHS_{it-1} - \bar{\phi}_a \frac{\gamma}{\alpha_M} k_{it-1} - \bar{\phi}_a \frac{\alpha_L}{\alpha_M}(l_{it-1} - k_{it-1}) \\ &+ \left(\bar{\phi}_\lambda - \bar{\phi}_a\right)\left(\frac{r_{it-1}}{s_{Mit-1}} - \frac{1}{\alpha_M} q_{it-1}\right) + \frac{1}{\alpha_M}\left(\nu_{ait} + \nu_{\lambda it}\right). \end{aligned} \tag{C-10}$$

where $\bar{\phi}_\lambda = \phi^*_\lambda + \phi^{'}_\lambda$ and $\bar{\phi}_a = \phi^*_a + \phi^{'}_a$. (C-10) is isomorphic to (21) and so the first stage of our baseline procedure can be applied to recover $\hat{b}_1$, $\hat{b}_2$ and $\hat{\bar{\phi}}_a$. However, the estimate of $\bar{\phi}_a$ is not useful anymore because in the second stage quantity equation we have instead ϕ^*_a (as well as $\phi_\lambda^{'}$):

$$\begin{aligned} q_{it} &= \frac{\gamma \hat{b}_2}{\hat{b}_1}(l_{it}-k_{it}) + \frac{\gamma}{\hat{b}_1}(m_{it}-k_{it}) + \gamma k_{it} + \phi_a^* \frac{\gamma}{\hat{b}_1} LHS_{it-1} \\ &- \phi_a^* \gamma k_{it-1} - \phi_a^* \frac{\gamma \hat{b}_2}{\hat{b}_1}(l_{it-1}-k_{it-1}) + (\phi_\lambda^{'} - \phi_a^*)\left(r_{it-1}\frac{\gamma}{\hat{b}_1 s_{Mit-1}} - q_{it-1}\right) + \nu_{ait} \end{aligned} \quad \text{(C-11)}$$

In order to recover γ one could finally rewrite (C-11) as the following linear regression:

$$q_{it} = b_8 z_{8it} + b_9 z_{9it} + b_{10} z_{10it} + b_{11} z_{11it} + \nu_{ait} \quad \text{(C-12)}$$

where:

$$z_{8it} = \frac{\hat{b}_2}{\hat{b}_1}(l_{it}-k_{it}) + \frac{1}{\hat{b}_1}(m_{it}-k_{it}) + k_{it}$$

$$z_{9it} = \frac{1}{\hat{b}_1} LHS_{it-1} - k_{it-1} - \frac{\hat{b}_2}{\hat{b}_1}(l_{it-1}-k_{it-1})$$

$$z_{10it} = \frac{r_{it-1}}{\hat{b}_1 s_{Mit-1}}$$

$$z_{11it} = q_{it-1}$$

as well as b_8=γ, b_9=$\gamma\phi_a^*$, b_{10}=$\gamma(\phi_\lambda^{'} - \phi_a^*)$ and b_{11}=$(\phi_a^* - \phi_\lambda^{'})$. Finally, z_{9it} to z_{11it} and predetermined and so uncorrelated with ν_{ait} while z_{8it} can, for example, be instrumented with capital at time t. This in turn allows recovering $\hat{\gamma}$ that, together with $\hat{b}_1$ and $\hat{b}_2$, can be used to recover productivity, product appeal and markups.

D Multi-product firms

There are several issues related to multi-product firms. We focus here on the issue of the assignment of inputs to outputs. Produced quantities and generated revenues may be observable for the different products of each firm in databases like ours. However, information on inputs used for a specific product is typically not available. We propose here an extension of our baseline model to solve the problem of assigning inputs to outputs for multi-product firms. In doing so we assume, as in De Loecker et al. (2016), there is a limited role for economies (or diseconomies) of scope on the cost side. However, contrary to De Loecker et al. (2016), we do not impose multi-product firms to be characterized by a common productivity across the different products they produce. We also allow for firm-product-time specific markups but impose product appeal/quality to be common across products within a firm. This corresponds to a setting where firms can be distinguished into those consistently selling high quality products and those consistently selling low quality products. Yet firms are allowed to be more or less efficient in the production of a specific product and charge different markups. The assumptions we lay down below and the related estimation procedure are consistent with both a monopolistically competitive market structure, like the one developed in Bernard et al. (2011), and the Cournot competition version of the model developed in Hottman et al. (2016) that we discuss in Appendix A.

As usual we denote a firm by i and time by t. A firm i produces in t one or more products indexed by p and the number of products produced by the firm is denoted by I_{it}. In our data p is an 8-digit prodcom product code but in other data, like the bar-code data used in Hottman et al. (2016), can be much more detailed. We assume product appeal is firm-time specific (λ_{it}) while we allow markups (μ_{ipt}) and productivity (a_{ipt}) to be firm-product-time specific. The production function for product p produced by firm i is given by:

$$Q_{ipt} = C_p C_t A_{ipt} L_{ipt}^{\alpha_{Lg}} M_{ipt}^{\alpha_{Mg}} K_{ipt}^{\gamma_g - \alpha_{Mg} - \alpha_{Lg}}, \tag{D-1}$$

where C_p and C_t are innocuous product and time constants we disregard in what follows and g identifies a product group/industry. Production function coefficients are the same for products within a product group because a certain level of data aggregation is needed to deliver enough observations to estimate parameters. (D-1) means we allow for technology (α_{Lg}, α_{Mg}, γ_g) to differ across the different products p produced by a multi-product firm. At the same time productivity is allowed to vary across products within a firm and information coming from single-product firms need to be used to infer the technology of multi-product firms, i.e., we rule out physical synergies in production but allow for some of the economies (diseconomies) of scope discussed in De Loecker et al. (2016). Furthermore, we assume firm i to maximize profits and choose (for each product p) the amount of labour L_{ipt} and materials M_{ipt} in order to minimize short-term costs while taking capital K_{ipt}, as well as productivity a_{ipt} and product appeal λ_{it} as given. We make use of (9) and so assume:

$$r_{ipt} \approx \frac{1}{\mu_{ipt}}(q_{ipt} + \lambda_{it}). \tag{D-2}$$

Profit maximization implies:

$$P_{ipt} = \mu_{ipt} \frac{\partial C_{ipt}}{\partial Q_{ipt}}, \tag{D-3}$$

where marginal cost is equal to[31]

$$\frac{\partial C_{ipt}}{\partial Q_{ipt}} = A_{ipt}^{-\frac{1}{\alpha_{Lg}+\alpha_{Mg}}} Q_{ipt}^{\frac{1-\alpha_{Lg}-\alpha_{Mg}}{\alpha_{Lg}+\alpha_{Mg}}} K_{ipt}^{\frac{\gamma_g-\alpha_{Lg}-\alpha_{Mg}}{\alpha_{Lg}+\alpha_{Mg}}} \tag{D-4}$$

Firms minimize costs and so markups are such that:

$$\mu_{ipt} = \frac{\alpha_{Mg}}{s_{Mipt}} \tag{D-5}$$

where s_{Mipt} is the expenditure share of materials for product p at time t in firm revenue for product p at time t. Finally, we assume that both a_{ipt} and λ_{it} evolve over time as linear stochastic Markov processes:

$$\begin{aligned} a_{ipt} &= \phi_{ag}\ a_{ipt-1} + \nu_{aipt} \\ \lambda_{it} &= \phi_{\lambda}\lambda_{it-1} + \nu_{\lambda it} \end{aligned}$$

where ν_{aipt} and $\nu_{\lambda it}$ can be correlated with each other.

As far as single-product firms are concerned the above assumptions are such that the parameters of the production function, as well as single-product firms productivity, product appeal and markups, can be obtained using a variant of the MULAMA procedure. More specifically, if also labour is chosen at t (21) becomes:

$$\begin{aligned} LHS_{ipt} &= \frac{\gamma_g}{\alpha_{Mg}} k_{ipt} + \phi_{ag} LHS_{ipt-1} - \phi_{ag}\frac{\gamma_g}{\alpha_{Mg}} k_{ipt-1} \\ &+ (\phi_{\lambda} - \phi_{ag})\left(\frac{r_{ipt-1}}{s_{Mipt-1}} - \frac{1}{\alpha_{Mg}} q_{ipt-1}\right) + \frac{1}{\alpha_{Mg}}\left(\nu_{aipt} + \nu_{\lambda it}\right), \end{aligned} \tag{D-6}$$

where $LHS_{ipt} \equiv \frac{r_{ipt} - s_{Lipt}(l_{ipt}-k_{ipt}) - s_{Mipt}(m_{ipt}-k_{ipt})}{s_{Mipt}}$ and s_{Lipt} is the share of labour in revenue ($s_{Lipt} \equiv \frac{W_{Lpt}L_{ipt}}{R_{ipt}}$). One can rewrite (D-6) as the following linear regression:

$$LHS_{ipt} = b_{1g}z_{1ipt} + b_{2g}z_{2ipt} + b_{3g}z_{3ipt} + b_{4g}z_{4ipt} + b_{5g}z_{5ipt} + u_{ipt} \tag{D-7}$$

where $z_{1ipt}{=}k_{ipt}$, $z_{2ipt}{=}LHS_{ipt-1}$, $z_{3ipt}{=}k_{ipt-1}$, $z_{4ipt}{=}\frac{r_{ipt-1}}{s_{Mipt-1}}$, $z_{5ipt}{=}q_{ipt-1}$, $u_{ipt}{=}\frac{1}{\alpha_{Mg}}\left(\nu_{aipt} + \nu_{\lambda it}\right)$ as well as $b_{1g}{=}\beta_g \equiv \frac{\gamma_g}{\alpha_{Mg}}$, $b_{2g}{=}\phi_{ag}$, $b_{3g}{=}-\phi_{ag}\beta_g$, $b_{4g}{=}(\phi_{\lambda} - \phi_{ag})$ and $b_{5g}{=}-\left(\phi_{\lambda} - \phi_{ag}\right)\frac{1}{\alpha_{Mg}}$. Given our assumptions, the error term u_{ipt} in (D-7) is uncorrelated with all of the regressors. Therefore (D-7) can be estimated via simple OLS. After doing this we set $\hat{\beta}_g{=}\hat{b}_{1g}$ and $\hat{\phi}_{ag}{=}\hat{b}_{2g}$ and do not exploit parameters' constraints in the estimation.

[31]We omit the innocuous product-time constant $\left(\frac{W_{Lpt}}{\alpha_{Lg}}\right)^{\frac{\alpha_{Lg}}{\alpha_{Lg}+\alpha_{Mg}}}\left(\frac{W_{Mpt}}{\alpha_{Mg}}\right)^{\frac{\alpha_{Mg}}{\alpha_{Lg}+\alpha_{Mg}}}$

We now turn to estimating γ_g. Within this setting we have $\alpha_{Lg} = \mu_{ipt} s_{Lipt}$ and $\alpha_{Mg} = \mu_{ipt} s_{Mipt}$. (25) becomes:

$$\begin{aligned} q_{ipt} &= \frac{\gamma_g}{\hat{\beta}_g} \frac{s_{Lipt}}{s_{Mipt}} (l_{ipt} - k_{ipt}) + \frac{\gamma_g}{\hat{\beta}_g} (m_{ipt} - k_{ipt}) + \gamma_g k_{ipt} \\ &+ \hat{\phi}_{ag} \frac{\gamma_g}{\hat{\beta}_g} LHS_{ipt-1} - \hat{\phi}_{ag} \gamma_g k_{ipt-1} - \hat{\phi}_{ag} \left(r_{ipt-1} \frac{\gamma_g}{\hat{\beta}_g s_{Mipt-1}} - q_{ipt-1} \right) + \nu_{aipt.} \end{aligned} \tag{D-8}$$

(D-8) can be rewritten in a linear way:

$$\overline{LHS}_{ipt} = b_{6g} z_{6ipt} + \nu_{aipt} \tag{D-9}$$

where:

$$\overline{LHS}_{ipt} = q_{ipt} - \hat{\phi}_{ag} q_{ipt-1}$$

$$z_{6ipt} = \frac{1}{\hat{\beta}_g} \frac{s_{Lipt}}{s_{Mipt}} (l_{ipt} - k_{ipt}) + \frac{1}{\hat{\beta}_g} (m_{ipt} - k_{ipt}) + k_{ipt} + \frac{\hat{\phi}_{ag}}{\hat{\beta}_g} LHS_{ipt-1} - \hat{\phi}_{ag} k_{ipt-1} - r_{ipt-1} \frac{\hat{\phi}_{ag}}{\hat{\beta} s_{Mipt-1}}$$

as well as $b_{6g}{=}\gamma_g$ and z_{6ipt} can instrumented with with k_{ipt} as well as past inputs, revenue and quantity. We set $\hat{\gamma}_g{=}\hat{b}_{6g}$ and are in turn able to estimate productivity as:

$$\hat{a}_{ipt} = q_{ipt} - \frac{\hat{\gamma}_g}{\hat{\beta}_g} \frac{s_{Lipt}}{s_{Mipt}} (l_{ipt} - k_{ipt}) - \frac{\hat{\gamma}_g}{\hat{\beta}_g} (m_{ipt} - k_{ipt}) - \hat{\gamma}_g k_{ipt}, \tag{D-10}$$

while product appeal and markups are computed as in the baseline procedure from (D-2) and (D-5).

Estimations need to be carried on single-product firms separately for each product group g. Turning to multi-product firms we impose, as in De Loecker et al. (2016), that the same technology parameters coming from single-product producers extend to the products of the former. Yet, in order to quantify multi-product firms productivity, markups and product appeal we still need to solve the issue of how to assign inputs to outputs and we do so by building on the above assumptions. As far as materials are concerned, we need to assign the observable total firm material expenditure M_{it} across the I_{it} products produced by firm i at time t, i.e., we need to assign values to M_{ipt} such that $\sum_{p=1}^{I_{it}} M_{ipt} = M_{it}$. We can use this condition along with (D-5) and (D-2) to operate this assignment. Substituting (D-5) into (D-2) and adding $\sum_{p=1}^{I_{it}} M_{ipt} = M_{it}$ provides a system of $I_{it} + 1$ equations in $I_{it} + 1$ unknowns; the I_{it} inputs expenditures M_{ipt} plus λ_{it}. Indeed, at this stage we have data on r_{ipt}, q_{ipt}, α_{Mg} and M_{it}. Operationally, one can actually proceed in two stage. Combining the above equations one has $\sum_{p=1}^{I_{it}} \frac{\alpha_{Mg} r_{ipt} R_{ipt}}{q_{ipt} + \lambda_{it}} = M_{it}$. This equation can be solved for each firm and delivers λ_{it}. With this at hand one can then obtain materials expenditure from

$M_{ipt} = \frac{\alpha_{Mg} r_{ipt} R_{ipt}}{q_{ipt}+\lambda_{it}}$. By recovering inputs expenditures M_{ipt} we can subsequently compute materials expenditure shares in revenues s_{Mipt} and so use (D-5) to recover our firm-product-time specific markups μ_{ipt}. Since labour is a variable input a condition analogous to (D-5) holds for this input and so we can use the computed markups μ_{ipt} and information on α_{Lg} to derive labour expenditure: $L_{ipt} = \frac{\alpha_{Lg} R_{ipt}}{\mu_{ipt}}$.[32] Operationally, this is not guaranteed to satisfy the constraint $\sum_{p=1}^{I_{it}} L_{ipt} = L_{it}$ for each firm and so the L_{ipt} need to be re-scaled for each firm.

The above procedure allows so far to obtain markups and product appeal, as well as information on labour and materials use, for each of the products of a multi-product firm. However, in order to recover productivity a_{ipt} we still need values for capital K_{ipt}. To do this one can proceed as follows. Combining the marginal cost, profit maximization and quantity equations one gets:

$$K_{ipt} = \left(\frac{P_{ipt}}{\mu_{ipt} Q_{ipt}^{a+b} L_{ipt}^{-a\alpha_{Lg}} M_{ipt}^{-a\alpha_{Mg}}} \right)^{\left(\frac{1}{c-a\alpha_{Kg}}\right)}$$

where $a = -\frac{1}{\alpha_{Lg}+\alpha_{Mg}}$, $b = \frac{1-\alpha_{Lg}-\alpha_{Mg}}{\alpha_{Lg}+\alpha_{Mg}}$, $c = \frac{\gamma_g-\alpha_{Lg}-\alpha_{Mg}}{\alpha_{Lg}+\alpha_{Mg}}$ and $\alpha_{Kg} = \gamma_g - \alpha_{Mg} - \alpha_{Lg}$ is the capital coefficient. Again values can be re-scaled for each firm to meet the constraint $\sum_{p=1}^{I_{it}} K_{ipt} = K_{it}$ and even further refined by running an estimation where the computed K_{ipt} is regressed on R_{ipt}, M_{ipt}, L_{ipt} as well as total firm expenditure on materials and labour plus the capital stock and product dummies.

E Conditional input demand for materials

To alleviate notation we derive here the conditional input demand for materials in the case where labour is a fully flexible input chosen in t as materials. The same steps can be used to derive the conditional input demand for materials in the situation where labour is predetermined or semi-flexible. This would lead to the same conclusion, i.e., invertibility does not generally apply because of the presence of firms with low markups.

The first step towards deriving the conditional input demand is to combine the marginal cost equation (E-1) with the expression for demand (E-2) and the profit maximization condition (E-3) given below.

$$\frac{\partial C_{it}}{\partial Q_{it}} = \frac{1}{\alpha_L + \alpha_M} \frac{C_{it}}{Q_{it}}. \tag{E-1}$$

[32]As a matter of fact in this variant of the model we do not impose α_{Lg} to be the same across firms. From every single product firm equation (D-5) applied to labour delivers, using the computed markups and the observed labour expenditure share in revenue, a different α_{Lg}. One can compute the mean value of these coefficients across firms producing products belonging to g to get a unique α_{Lg}.

$$Q_{it} \approx P_{it}^{\frac{\mu_{it}}{1-\mu_{it}}} \Lambda_{it}^{\frac{1}{1-\mu_{it}}}. \tag{E-2}$$

$$P_{it} = \mu_{it} \frac{\partial C_{it}}{\partial Q_{it}}. \tag{E-3}$$

Solving for the optimal price delivers:

$$P_{it}^{\frac{\mu_{it}}{\mu_{it}-1}\left(\frac{1-\alpha_L-\alpha_M}{\alpha_L+\alpha_M}\right)+1} = \mu_{it} K_{it}^{-\frac{\gamma-\alpha_L-\alpha_M}{\alpha_L+\alpha_M}} A_{it}^{-\frac{1}{\alpha_L+\alpha_M}} \Lambda_{it}^{\frac{1}{1-\mu_{it}}\left(-\frac{1-\alpha_L-\alpha_M}{\alpha_L+\alpha_M}\right)} \kappa^{\frac{\mu_{it}}{\mu_{it}-1}\left(-\frac{1-\alpha_L-\alpha_M}{\alpha_L+\alpha_M}\right)} C,$$

where $C = \left(\frac{W_L}{\alpha_L}\right)^{\frac{\alpha_L}{\alpha_L+\alpha_M}} \left(\frac{W_M}{\alpha_M}\right)^{\frac{\alpha_M}{\alpha_L+\alpha_M}}$ is constant across firms. The optimal price is thus a function of capital, productivity, product appeal and the markup. Substituting the optimal price into the expression for demand (E-2) delivers the equilibrium quantity and multiplying this by the optimal price gives equilibrium firm revenue:

$$R_{it} = P_{it} Q_{it} = \mu_{it}^{\frac{1}{-\mu_{it}\left(\frac{1-\alpha_L-\alpha_M}{\alpha_L+\alpha_M}\right)+(1-\mu_{it})}} K_{it}^{\frac{-\frac{\gamma-\alpha_L-\alpha_M}{\alpha_L+\alpha_M}}{-\mu_{it}\left(\frac{1-\alpha_L-\alpha_M}{\alpha_L+\alpha_M}\right)+(1-\mu_{it})}} A_{it}^{\frac{-\frac{1}{\alpha_L+\alpha_M}}{-\mu_{it}\left(\frac{1-\alpha_L-\alpha_M}{\alpha_L+\alpha_M}\right)+(1-\mu_{it})}}$$
$$\Lambda_{it}^{\frac{-\frac{1-\alpha_L-\alpha_M}{\alpha_L+\alpha_M}}{-\mu_{it}(1-\mu_{it})\left(\frac{1-\alpha_L-\alpha_M}{\alpha_L+\alpha_M}\right)+(1-\mu_{it})^2}-\frac{1}{1-\mu_{it}}} \kappa^{\frac{-\frac{1-\alpha_L-\alpha_M}{\alpha_L+\alpha_M}}{(1-\mu_{it})\left(\frac{1-\alpha_L-\alpha_M}{\alpha_L+\alpha_M}\right)+\frac{(1-\mu_{it})(\mu_{it}-1)}{\mu_{it}}}-\frac{\mu_{it}}{1-\mu_{it}}} C^{\frac{1}{-\mu_{it}\left(\frac{1-\alpha_L-\alpha_M}{\alpha_L+\alpha_M}\right)+(1-\mu_{it})}}, \tag{E-4}$$

which is again a function of capital, productivity, product appeal and the markup.

From (5) the expenditure in materials is simply $W_M M_{it} = \frac{\alpha_M R_{it}}{\mu_{it}}$ and so the elasticity of $W_M M_{it}$ with respect to A_{it} is $\frac{-\frac{1}{\alpha_L+\alpha_M}}{-\mu_{it}\left(\frac{1-\alpha_L-\alpha_M}{\alpha_L+\alpha_M}\right)+(1-\mu_{it})}$. Such elasticity is positive (negative) as long as μ_{it} is larger (smaller) than $\alpha_L + \alpha_M$. Therefore, invertibility does not generally apply because of the presence of firms with low markups (below $\alpha_L + \alpha_M$). Also note the above condition transforms into μ_{it} being larger (smaller) than α_M in the case where labour is either predetermined or semi-flexible.

Firms with low markups are present in our data as well as in other datasets and analyses. For example, given an average markup in between 1.1 and 1.2 and a standard deviation of 0.5 quite a few firms in De Loecker and Warzynski (2012) would fall into this category. Yet, this is not an issue within our estimation procedure because we do not need invertibility.

F Additional Tables

Table F-1: Estimates of production function parameters with the DLG-WLD procedure

Industry	Description	Labour	Materials	Capital
1	Food products, beverages and tobacco	0.276^a	0.580^a	0.033^b
		(0.063)	(0.160)	(0.015)
2	Textiles and leather	0.255^a	0.544^a	0.035^a
		(0.043)	(0.120)	(0.007)
3	Wood except furniture	0.235^a	0.608^a	0.004
		(0.059)	(0.093)	(0.007)
4	Pulp, paper, publishing and printing	0.249^a	0.656^a	0.012^b
		(0.030)	(0.061)	(0.006)
5	Chemicals and rubber	0.123^a	0.886^a	0.015^c
		(0.029)	(0.052)	(0.008)
6	Other non-metallic mineral products	0.151^a	0.814^a	0.007
		(0.022)	(0.034)	(0.005)
7	Basic metals and fabric. metal prod.	0.271^a	0.648^a	0.026^a
		(0.017)	(0.020)	(0.004)
8	Machinery, electric. and optical equip.	0.288^a	0.620^a	0.027^a
		(0.032)	(0.031)	(0.004)
9	Transport equipment and n.e.c.	0.036	0.852^a	0.024^a
		(0.064)	(0.056)	(0.008)

Notes: Robust standard errors in parenthesis. ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1.

Table F-2: Regression of markups on TFP, product appeal (λ) and log capital by industry

Industry	1	2	3	4	5	6	7	8	9
TFP	0.2024^a	0.2476^b	0.4859^a	0.2911^a	0.2258^a	0.2287^a	0.5354^a	0.3942^a	0.3427^a
	(0.0415)	(0.0925)	(0.1014)	(0.0525)	(0.0305)	(0.0421)	(0.0374)	(0.0478)	(0.0854)
λ	0.2879^a	0.2786^b	0.5172^a	0.3067^a	0.2426^a	0.2886^a	0.5322^a	0.4063^a	0.3521^a
	(0.0391)	(0.0854)	(0.0949)	(0.0476)	(0.0293)	(0.0341)	(0.0363)	(0.0479)	(0.084)
capital	-0.0472^a	-0.0199^c	0.0056	-0.0266^b	-0.0154^c	-0.0176^b	-0.0223^a	-0.0261^c	-0.028^b
	(0.0109)	(0.0088)	(0.0177)	(0.0089)	(0.0066)	(0.0067)	(0.0056)	(0.0102)	(0.0086)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Prod dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N Obs	1317	1225	348	975	1043	1215	2814	1108	1055
R^2	0.6076	0.6265	0.609	0.6078	0.796	0.7749	0.6036	0.7201	0.4372

Notes: The dependent variable is the markup. Bootstrapped standard errors in parenthesis (200 replications). ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1.

Table F-3: Regression of demeaned log prices on demeaned TFP, product appeal (λ) and log capital by industry

Industry	1	2	3	4	5	6	7	8	9
TFP	-0.8336^a	-1.0010^a	-0.9682^a	-0.9245^a	-0.8223^a	-0.9048^a	-0.9502^a	-0.986^a	-0.7973^a
	(0.0454)	(0.022)	(0.0537)	(0.0441)	(0.0454)	(0.0178)	(0.02)	(0.0201)	(0.0547)
λ	0.0977^b	-0.0108	0.0237	0.0473	0.1656^a	0.0606^a	0.0415^c	0.0059	0.1928^a
	(0.0375)	(0.0206)	(0.051)	(0.0385)	(0.0437)	(0.0135)	(0.019)	(0.0191)	(0.0527)
capital	-0.1096^a	-0.0068^c	-0.0099	0.0442^a	0.0098	-0.01^c	-0.0060	-0.0069^c	0.0051
	(0.0075)	(0.003)	(0.0093)	(0.0092)	(0.0096)	(0.0039)	(0.0027)	(0.0032)	(0.004)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N Obs	1317	1225	348	975	1043	1215	2814	1108	1055
R^2	0.9278	0.9933	0.9937	0.9799	0.9938	0.9824	0.9958	0.9979	0.9948

Notes: Prices, TFP, product appeal and capital are demeaned by 8-digit Prodcom codes. The dependent variable is the demeaned log price. Bootstrapped standard errors in parenthesis (200 replications). ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1.

Table F-4: Fixed effects regression (within estimator) of TFP, product appeal (λ) and markups on their time lag by industry

Industry									
					TFP				
lag TFP	0.5925^a	0.6678^a	0.5844^a	0.6804^a	0.7553^a	0.5804^a	0.4838^a	0.3737^b	0.3236^b
	(0.0552)	(0.053)	(0.1681)	(0.1325)	(0.0378)	(0.1228)	(0.1104)	(0.1176)	(0.1222)
N Obs	901	843	232	710	702	867	2000	785	738
R^2	0.3805	0.3683	0.438	0.3893	0.6116	0.3684	0.2312	0.1868	0.1367
					λ				
lag λ	0.6246^a	0.6667^a	0.5696^a	0.6022^a	0.7492^a	0.5079^a	0.4888^a	0.3741^a	0.315^c
	(0.039)	(0.049)	(0.1266)	(0.1819)	(0.0384)	(0.1023)	(0.0949)	(0.1028)	(0.1313)
N Obs	901	843	232	710	702	867	2000	785	738
R^2	0.3814	0.3749	0.4001	0.26	0.5931	0.2487	0.2309	0.184	0.1359
					markup				
lag markup	0.5132^a	0.3858^a	0.605^a	0.3654^a	0.5415^a	0.4382^a	0.3435^a	0.3007^a	0.3083^a
	(0.057)	(0.0506)	(0.1022)	(0.0839)	(0.0729)	(0.0777)	(0.0354)	(0.0512)	(0.0771)
N Obs	901	843	232	710	702	867	2000	785	738
R^2	0.3286	0.1771	0.4255	0.2332	0.4553	0.3404	0.1834	0.1617	0.1862

Notes: Time dummies are included in estimations but are not reported here. Bootstrapped standard errors in parenthesis (200 replications). ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1.