Identifying Unobserved Heterogeneity in Productivity

Ruben Dewitte, Catherine Fuss and Angelos Theodorakopoulos*

30th April 2021

Abstract

Maintaining economic prosperity in an aging economy like Belgium will require consistently strong aggregate productivity growth. A correct identification of the micro-economic engines of this productivity growth will be essential for policy guidance, but current identification strategies are hampered by heavy data requirements. Whereas the evolution of firm-level productivity is known to be influenced by firm-level characteristics such as innovation, management practices, trade ..., these characteristics are often unavailable to the researcher. Productivity measurements and the identification of their main determinants therefore suffer from an omitted variable bias, as unobserved heterogeneity is left uncaptured. This papers uncovers and aims to resolve this omnipresent bias in the measurement of firm-level productivity. It proposes an identification methodology that aims to capture unobserved heterogeneity in productivity using a flexible, semi-parametric extension (Finite Mixture Models) of state-of-the-art estimation techniques. A Monte Carlo analysis demonstrates how the proposed estimation methodology is able to correct the bias in current production function parameter estimates, while an application to Belgian firm-level data demonstrates the applicability of the proposed methodology.

Keywords: Finite Mixture Model, firm size distribution, productivity distribution

JEL Codes: L11

*Dewitte (corresponding author): Ghent University, Tweekerkenstraat 2, 9000 Ghent, Belgium, rubenl.dewitte@ugent.be, Fuss: National Bank of Belgium, boulevard de Berlaimont 14, 1000 Brussels, Catherine.Fuss@nbb.be , Theodorakopoulos: Oxford Martin School, University of Oxford, 34 Broad Street, Oxford OX1 3BD, United Kingdom, angelos.theodorakopoulos@oxfordmartin.ox.ac.uk.

1 Introduction

Maintaining economic prosperity in an aging economy as Belgium will require consistently strong productivity growth. According to the European Ageing Working Group, social security costs will increase with ±5.0–7.3% of Gross Domestic Product by 2070 in Belgium (European Commission, 2015). To curb these costs, it has been posited aggregate productivity should grow by ±1.2% on average yearly, even though productivity growth has been consistently lower ever since the global economic and financial crisis of 2008-09 (Peersman, 2019).

While a correct identification of the micro-economic engines of productivity growth will be essential for policy guidance, current identification strategies are hampered by heavy data requirements. Whereas the evolution of firm-level productivity is known to be differ between clusters of firms driven by firm-level characteristics such as innovation (Costantini and Melitz, 2008; Bee et al., 2011; Atkeson and Burstein, 2010), management practices (Caliendo et al., 2015; Bloom and Van Reenen, 2011), trade (Kasahara and Rodrigue, 2008; De Loecker, 2013; Kasahara and Lapham, 2013), industry linkages (Luttmer, 2007) ..., these characteristics are often unavailable to the researcher. Productivity measurements and the identification of their main determinants therefore often suffer from an omitted variable bias, as unobserved heterogeneity in the productivity growth process is left uncaptured.

This paper uncovers and aims to resolve the omnipresent omitted variable bias in the measurement of firm-level productivity. The proposal builds on the observation of **?** that firm productivity is currently captured assuming a homogeneous random growth process for all firms (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015; Gandhi et al., 2020; Forlani et al., 2016). If heterogeneity from firm-level characteristics is present but not controlled for, productivity estimates will be biased, imposing huge data requirements for current productivity estimation methodologies. This project proposes an extension to current productivity estimation methodologies using Finite Mixture Models (FMM). FMMs allow for the productivity evolution to differ between clusters of firms even though the drivers of these differences are 'unobserved'.

The proposed methodology builds on the behavioral framework set out by (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015; Gandhi et al., 2020; Forlani et al., 2016). The firm's production output is a function of factor inputs and Hicks-neutral productivity. Rather than specifying this productivity as the outcome of a growth process common to all firms, we allow for productivity to evolve differently between clusters of firms. At firm entry, the firm makes a one-off decision on this cluster affiliation depending on its initial and expected future ability to produce output from factor inputs and a, to the researcher unknown, affinity for certain clusters. This unobserved affinity can be driven by industry affiliation, the innovative ability of the firm, the potential to trade, the ideology regarding management practices ...

We build on a distributional assumption regarding the firm's affinity for productivity cluster to model the probability of cluster affiliation for each firm and cluster. Our work demonstrates that the parameters of the production function are identified based on the independence of timing decisions of factor input choice and *cluster-probability weighted* shocks to productivity (see (Gandhi et al., 2020) for a recent overview of current prevalent identification schemes). To control for the simultaneity of these factor input decisions in this semi-parametrically defined environment, we rely on Limited Information Maximum Likelihood (LIML) rather than the prevalent non-parametric Generalized Method of Moments (GMM). We use the Expectation-Maximization (EM) algorithm to estimate the production function parameters and simultaneously identify a firm's cluster affiliation. Thus, by allowing for unobserved heterogeneity in productivity we can obtain 'unbiased'[1] estimates of productivity while greatly reducing data requirements.

We demonstrate the appropriateness of the proposed methodology in an Monte Carlo analysis and the applicability in an application to Belgian firm-level data. First, we extend the Monte Carlo exercise of (Ackerberg et al., 2015) to account for unobserved heterogeneity in productivity. We demonstrate how current estimation methodologies produce biased estimates of the production function when such heterogeneity in the evolution of productivity is present, and how the our methodology is able to correct for this bias and simultaneously identify firm cluster affiliation. An application to Belgian firm-level data confirms the existence of such biases. We rely on a vast dataset of firm-level characteristics (location, age, size, industry affiliation, FDI status, trade status, ...) to reveal economy-relevant differences in the evolution of productivity between clusters of firms.

The rest of the paper is as follows. In section 2, we embed the paper in the current scientific literature before exposing and testing the proposed methodology in section 3. We apply the methodology to Belgian firm-level data in section 4 and finish by summarizing the main contributions and future research directions in section 5.

2 Literature Review

This paper mainly builds on the structural production function estimation literature. Provided certain assumption regarding (i) the functional relation between output and inputs, (ii) the timing of input decisions and (iii) the evolution of productivity, this literature aims to identify unobserved (to the researcher) productivity from a production function estimation (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015; Gandhi et al., 2020; Forlani et al., 2016). The focus in this project lies on the prevalent assumption of the evolution of productivity of firm b at time t, ω_{bt}, which is assumed to follow a first-order Markov process:

$$\omega_{bt} = g(\omega_{bt-1}) + \eta_{bt}, \tag{1}$$

where $g(\cdot)$ is a currently unspecified functional form and η_{bt} represents an identically and independently distributed (i.i.d.) error term.

The predominant specification displayed in 1 assumes the evolution of productivity to be a function of its lagged values and a random error term: a homogeneous random growth process

[1] We refer to unbiased estimates as estimates having resolved the uncovered omitted variable bias, without making a statement on the appropriateness of the methodology to correct other possible biases occurring during the estimation of productivity. See Van Beveren (2012) for an overview.

for all firms. Any determinative heterogeneity, in function of certain firm-level characteristics represented by the vector $\boldsymbol{e}_{bt}$, that results in a different growth process between clusters of firms needs to be specified:

$$\omega_{bt} = \tilde{g}(\omega_{bt-1}, \boldsymbol{e}_{bt}) + \eta_{bt}. \tag{2}$$

Such specification allows for idiosyncratic evidence on the endogenous evolution of firm-level productivity to firm-level characteristics such as innovation (Bee et al., 2011), management practices (Caliendo et al., 2015; Bloom and Van Reenen, 2011), trade (Kasahara and Rodrigue, 2008; De Loecker, 2013), human capital Van Beveren and Vanormelingen (2014) ...(see Table 1 for a non-exhaustive overview of the literature relying on such identification strategy).

Table 1: Non-exhaustive literature list identifying productivity drivers from within the Markov process.

Study	Export	Import	R&D	FDI	Others (industry, location, ...)
Olley and Pakes (1996)					Age, telecommunications industry
Javorcik (2004)				x	Manufacturing (plant-ind-location-time FE)
Amiti and Konings (2007)		x			Manufacturing
Das et al. (2007)	x				2-digit industry
Blalock and Gertler (2008)				x	Manufacturing (ind-location-time FE)
Kasahara and Rodrigue (2008)		x			Manufacturing
Aw et al. (2011)	x		x		Electronics industry
De Loecker (2013)	x				2-digit industry, investment
Doraszelski and Jaumandreu (2013)			x		2-digit industry, investment
Kasahara and Lapham (2013)	x	x			3-and 4- digit industry

Admitting heterogeneity as specified in equation 2 entails that estimating the production function without controlling for this heterogeneity, the firm-level drivers of productivity ($\boldsymbol{e}_{bt}$), results in an *omitted variable bias*. As already argued by De Loecker (2013), one cannot correctly identify the drivers of firm-level productivity if all relevant firm-level characteristics are not controlled for. This places heavy, if not impossible, data requirement on the estimation procedure, leading to biased production function and resulting productivity estimates.

This paper therefore proposes to extend current methodologies using Finite Mixture Modeling. FMMs allow us to identify clusters of firms with a homogeneous growth process

$$\omega_{bt} = \sum_{s=1}^{S} Pr(z_b^s|\cdot)\left[g^s(\omega_{bt-1}) + \eta_{bt}^s\right], \tag{3}$$

where the number of clusters S and the probability a firm i belongs to a certain cluster $s \in S$, $Pr(z_b^s|\cdot)$, are determined by the data. Differences between these firm clusters, even though determined by firm-level characteristics, can therefore be left unobserved.

We are not the first to explore the advantages of FMMs for the estimation of firm-level productivity. Closest related to our work are van Biesebroeck (2003), Kasahara et al. (2017) and Battisti et al. (2020). van Biesebroeck (2003) specifies non-Hicks neutral technology that differs between two clusters, with the possibility of transition from the lean to mass technology cluster, for a panel of U.S. automobile assembly plants. The probability of cluster affiliation is modelled as a function of the relative profitability of the lean vs. mass technology in reduce form. To control for the simultaneity problem when estimating the value-added production function, van Biesebroeck (2003) relies on a parametrization of the first-order conditions with respect to factor inputs. Similarly (Kasahara et al., 2017; Battisti et al., 2020,?) specify non-Hicks neutral technology with an a priori undefined number of clusters without the possibility of transition. Also here, the authors rely rely on a parametrization of the first-order conditions with respect to factor inputs to control for simultaneity when estimating the value-added (Battisti et al., 2020) or revenue (Kasahara et al., 2017) production function. This paper focuses on generalizing prevalent Hicks-neutral productivity estimation strategies using both value-added (Ackerberg et al., 2015) and revenue (Gandhi et al., 2020) production functions. We control for simultaneity problems relying on a Limited Information Maximum Likelihood specification, preventing us from imposing additional assumptions regarding the first-order conditions of factor inputs. Moreover, we model the probability of belonging to a certain cluster in line with the behavioral framework. This results in a so-called Mixture-of-experts specification (Gormley and Frühwirth-Schnatter, 2019) that improves cluster identification as well as allows for an evaluation of the economic relevance of the identified clusters.

The advantages of capturing productivity has also already been explored in the stochastic frontier literature (see for instance Beard et al. (1997); Greene (2005); Orea and Kumbhakar (2004); El-Gamal and Inanoglu (2005)). Compared to this literature, less stringent functional form restrictions are necessary for the proposed structural production function estimation techniques relied upon in this paper (Sickles and Zelenyuk, 2019). We do build on the idea, advanced in the stochastic frontier literature, of controlling for the simultaneity problem when estimating a production function using LIML (see (Amsler et al., 2016) for an overview).

3 Methodology

This section describes in detail the behavioral framework imposed on the data and subsequent production function identification and estimation strategy, before evaluating the proposed technique in a Monte Carlo exercise.

3.1 Behavioral framework

We assume a dynamic heterogeneous firms model with cluster-dependent uncertainty in future, Hicks-neutral, productivity. The data consists of a (short) panel of firms $b = 1, \ldots, B$ over period $t = 0, \ldots, T$. These firms produce a certain output Y_{bt} provided a certain amount of capital K_{bt}, labor L_{bt} and materials M_{bt} in perfectly competitive output and input markets.[2] The firms have access to information at time t, referred to as the information set $\mathcal{I}_{bt}$. This information set $\mathcal{I}_{bt}$ is available to the firm when making its production decisions in period t.

[2]While we limit the outline of the behavioral framework to perfect competition for simplicity, the proposed identification procedure solely affects the Markov assumption of productivity and can therefore naturally be extended to imperfect competition.

For a generic input $X_{bt} \in \{K_{bt}, L_{bt}, M_{bt}\}$, we say that this input is *non-flexible* if it is either predetermined $X_{bt} \in \mathcal{I}_{bt}$ or dynamic $X_{bt} = f(X_{bt-1})$. Similarly, it is a *flexible* input if it is neither predetermined $X_{bt} \notin \mathcal{I}_{bt}$ nor dynamic $X_{bt} \neq f(X_{bt-1})$ (Gandhi et al., 2017). In this paper, we assume that capital and labor are non-flexible while materials is a flexible input.

The relationship between outputs and inputs is of the form:

$$\begin{aligned} Y_{bt} &= F^{klm}(K_{bt}, L_{bt}, M_{bt})\, e^{\omega_{bt}+\epsilon_{bt}} \qquad \Leftrightarrow \\ y_{bt} &= f^{klm}(k_{bt}, l_{bt}, m_{bt}) + \omega_{bt} + \epsilon_{bt}, \end{aligned} \tag{4}$$

where log values of variables are written in lowercase. $f^{klm}(\cdot)$ represents the production function that explains the variability in firm-level output, next to two Hicks-neutral productivity shocks. $\epsilon_{bt} \notin \mathcal{I}_{bt}$ represents an ex-post Hicks-neutral productivity shock that does not affect future output. $\omega_{bt} \in \mathcal{I}_{bt}$, on the other hand, represents Hicks-neutral productivity that is known to the firm before making its period t decisions. Furthermore, this productivity component ω_{bt} evolves over time according to a *cluster-dependent* first-order Markov process

$$p(\omega_{bt}|\mathcal{I}_{bt-1}) = p(\omega_{bt}|\omega_{bt-1}, z_b^s). \tag{5}$$

Each firm b belongs to a certain cluster $s = 1, \ldots, S$, indicated by $z_b^i = \mathbb{I}_b(s = i), \forall i = 1, \ldots, S$, where $\mathbb{I}$ is the indicator function. $p(\cdot)$ indicates the probability density function of continuous variables.

When entering, the firm makes a net present value comparison between clusters and chooses the cluster to belong to from the next period onwards that results in the highest discounted profits, taking expectations and the costs of cluster affiliation into account. This results in an optimal decision rule:[3]

$$\begin{aligned} z_b^*(k_{b0}, l_{b0}, \omega_{b0}) = \max_{z_b^s} \Bigg(& \pi_{b0}(k_{b0}, l_{b0}, \omega_{b0}) + \epsilon(z_b^s) + \\ & E\left[\sum_{t=1}^{T} \beta^{t-1} \pi_{bt}(k_{bt}, l_{bt}, \omega_{bt}, z_b^s))\right] \Bigg) \end{aligned} \tag{6}$$

where $\beta \in \{0, 1\}$ is the discount factor, $\pi_{bt}(\cdot)$ a firm's profit and $\epsilon(z_b^s)$ is a choice-specific i.i.d. variable that captures the affinity of a firm for a certain cluster. Integrating out this affinity provides us with the probability of belonging to a certain cluster from period 1 onwards, conditional on initial capital, labor, productivity:

$$Pr(z_b^s|k_{b0}, l_{b0}, \omega_{b0}) = \int \mathbb{I}\left[z_b^*(k_{b0}, l_{b0}, \omega_{b0}) = z_b^s\right] g(\epsilon) d\epsilon. \tag{7}$$

[3] The firm has a certain productivity at entry and chooses cluster affiliation in its first period, depending on this productivity, that then determines the evolution of productivity from this period onwards.

3.2 Identification of the production function

An immediate identification of the production function parameters based on the production function specified in equation 4 is burdened by the simultaneity problem. Firm-level input choices on unobservable Hicks-neutral productivity, $E\left[\left(\omega_{bt}+\epsilon_{bt}\right)|k_{bt},l_{bt},m_{bt}\right]\neq 0$. This dependence renders Ordinary- (OLS) or Nonlinear- (NLLS) Least Squares estimates of the output elasticities inconsistent. Alternative identification strategies have been developed, which usually consist of two stages.

In a *first stage*, one tries to separate the deviations (ϵ_{bt}) from the main estimating equation (Eq. 4). Different methods exist to achieve this, without consensus on the superiority of any of these methods. For instance, Gandhi et al. (2020) build on the first-order conditions of the flexible production factors, (m_{bt}), to identify these deviations and the output elasticity of the flexible production factors simultaneously. Ackerberg et al. (2015), on the other hand, assume proportionality of the flexible production factors (m_{bt}) to output. They then rely on these production factors as control variables for productivity in order to separate out the measurement error. Disregarding the methodology relied upon, the first stage results in an equation of the form:

$$\phi_{bt}=f^{kl}\left(k_{bt},l_{bt}\right)+\omega_{bt}, \tag{8}$$

where $\hat{\phi}_{bt}$ represents non-flexible output variation, i.e. what is left of output variation once measurement error and the contribution of the flexible production factors to firm-level output are subtracted.

The *second stage*, then, focuses on the identification of the output elasticities of non-flexible inputs. It relies on the Markov property of productivity (see eq. 5) to replace the unobserved productivity term ω_{bt} in the previous equation (Eq. 8) as a function of observables and the assumed evolution of this productivity over time. As this evolution depends on the cluster affiliation, this paper additionally accounts for the conditional probability of belonging to a certain cluster (see eq. 7):

$$\phi_{bt}=f^{kl}\left(k_{bt},l_{bt}\right)+\sum_{s=1}^{S}Pr(z_b^s|k_{b0},l_{b0},\omega_{b0})\left[g^s\left(\phi_{bt-1}-f^{kl}\left(k_{bt-1},l_{bt-1}\right)\right)+\eta_{bt}^s\right]. \tag{9}$$

Identification of the production function parameters is obtained based on the independence of *probability-weighted*, cluster-specific expected productivity deviations (η_{bt}^s) from the current capital stock (k_{bt}), lagged non-flexible output variation ϕ_{bt-1} and, depending on the timing assumption of labor input decisions, either current (l_{bt}) or lagged labor l_{bt-1} (Arcidiacono and Jones, 2003; Ackerberg et al., 2015).

$$E\left[\sum_{s=1}^{S}Pr(z_b^s|k_{b0},l_{b0},\omega_{b0})\eta_{bt}^s\middle|k_{bt},l_{bt(-1)},\phi_{bt-1}\right]=0. \tag{10}$$

At this point, it is informative to discuss the possible parameterization for the conditional probability of cluster affiliation that allow identification. We consider (i) unitary probabilities and (ii) deterministic probabilities based on the literature, while proposing (iii) random probabilities as an alternative parameterization.

(i) Unitary probabilities If one assumes $S = 1$, then $Pr(z_b^s | k_{b0}, l_{b0}, \omega_{b0}) = 1$ and equation 9 takes the well-known form

$$\phi_{bt} = f^{kl}\left(k_{bt}, l_{bt}\right) + g^1\left(\phi_{bt-1} - f^{kl}\left(k_{bt-1}, l_{bt-1}\right)\right) + \eta_{bt}^1.$$

While this is the specification commonly relied upon in the literature, this specification results in an omitted variable bias if there is a least one group of firms that evolves differently over time, $S > 1$:[4]

$$E\left[\sum_{s=1}^{S} Pr(z_b^s | k_{b0}, l_{b0}, \omega_{b0}) \eta_{bt}^s \middle| k_{bt}, l_{bt(-1)}, \phi_{bt-1}\right] = 0 \neq E\left[\sum_{s=1}^{S} \eta_{bt}^1 \middle| k_{bt}, l_{bt(-1)}, \phi_{bt-1}\right].$$

(ii) Deterministic probabilities Assume we have access to an $N-$dimensional vector of categorical variables $\boldsymbol{e}_b = \left\{e_b^1, \ldots, e_b^N\right\}$ that determine cluster affiliation. If $N = S$, we have

$$Pr(z_b^s | k_{b0}, l_{b0}, \omega_{b0}) = \mathbb{I}\left(\boldsymbol{e}_b = e_b^i\right), \qquad \forall i = 1, \ldots, S, \tag{14}$$

and, as a result,

$$E\left[\sum_{s=1}^{S} \mathbb{I}\left(\boldsymbol{e}_b = e_b^s\right) \eta_{bt}^s \middle| k_{bt}, l_{bt(-1)}, \phi_{bt-1}\right] = 0. \tag{15}$$

Usually, to the researcher, it is priori unknown whether these variables are determinative of cluster affiliation, and whether N is larger, smaller or equal to S. They are proxy variables. If $N < S$, we again face an omitted variable bias. The likelihood of facing an omitted variable bias is present, as was emphasized in the literature discussion in section 2 and accompanying Table 1. Moreover, measurement error in the proxy variables will result in cluster misallocation with a biased estimator as a result.

[4]Assume Cobb-Douglas production function, AR(1) productivity and two clusters of firms:

$$\phi_{bt} = \beta_k k_{bt} + \beta_l l_{bt} + \mathbb{I}_b\left(s = 1\right)\left(\alpha_0^1 + \alpha_1^1\left(\phi_{bt-1} - \beta_k k_{bt-1} - \beta_l l_{bt-1}\right) + \eta_{bt}^1\right) \tag{11}$$

$$+ \mathbb{I}_b\left(s = 2\right)\left(\alpha_0^2 + \alpha_1^2\left(\phi_{bt-1} - \beta_k k_{bt-1} - \beta_l l_{bt-1}\right) + \eta_{bt}^2 \cdot\right). \tag{12}$$

When imposing a unitary prior, the specification becomes

$$\phi_{bt} = \beta_k k_{bt} + \beta_l l_{bt} + \alpha_0^* + \alpha_1^*\left(\phi_{bt-1} - \beta_k k_{bt-1} - \beta_l l_{bt-1}\right) + \eta_{bt}. \tag{13}$$

If $\alpha_{0,1}^{1,2} \gtrless 0$, by definition, the omitted cluster-indicator is correlated with the remaining explanatory variables and will positively/negatively bias the estimated coefficients (see, for instance, De Loecker (2013)).

(iii) Random probabilities Provided the risk for omitted variable bias, this paper relies on the behavioral framework presented in section 3.1. and a parametric assumption on the firms affinity towards certain clusters to specify cluster probabilities without additional information. Assuming the choice-specific i.i.d. variable $\epsilon(z_b^s)$ follows a type-1 extreme value distribution and relying on a reduced form optimal decision rule $(z^*(k_{b0}, l_{b0}, \omega_{b0}))$, we can model the conditional choice probability as (McFadden, 1973) :[5]

$$Pr(z_b^s|k_{b0}, l_{b0}, \omega_{b0}; \boldsymbol{\gamma}^1, \ldots, \boldsymbol{\gamma}^s) = \frac{e^{\gamma_0^i + \gamma_k^i k_{b0} + \gamma_l^i l_{b0} + \gamma_\omega^i \omega_{b0}}}{\sum_{s=1}^S e^{\gamma_0^s + \gamma_k^s k_{b0} + \gamma_l^s l_{b0} + \gamma_\omega^s \omega_{b0}}}, \qquad \forall i = 1, \ldots, S. \tag{16}$$

This approach has the advantage it allows to identify the parameters of the production function without prior knowledge on firm cluster affiliation:

$$E\left[Pr(z_b^s|k_{b0}, l_{b0}, \omega_{b0}; \boldsymbol{\gamma}^1, \ldots, \boldsymbol{\gamma}^s)\eta_{bt}^s \middle| k_{bt}, l_{bt(-1)}, \phi_{bt-1}\right] = 0. \tag{17}$$

Additionally, if prior information regarding cluster affiliation ($\boldsymbol{e}_b$) is available, the proposed approach is as good as (for $N = S$) or better than (for $N < S$) the deterministic approach.[6] Lastly, this approach even allows for imperfect proxy variables.

3.3 Estimation of the production function

Having established the production function parameters can be identified despite the presence unobserved heterogeneity in the evolution of productivity, we need to determine the estimation procedure for this production function. At this point, we focus the outline of the estimation methodology starting from the second stage estimation equation (see 9). We rely on a Finite Mixture specification for the evolution of productivity and the Expectation-Maximization algorithm to jointly estimate the production function parameters and the firm cluster affiliation. To alleviate possible endogeneity concerns regarding labor input (Ackerberg et al., 2015), we rely on a Limited Information Maximum Likelihood to allow for current labor to be instrumented with lagged labor values. We start by specifying the observed likelihood, conditioning on lagged labor values. We then specify the complete likelihood that accounts for cluster affiliation, before presenting the estimation algorithm.

[5] We rely on a reduced form optimal decision rule for simplicity. For a discussion on the estimation of dynamic discrete choice models with unobserved heterogeneity, we kindly refer the reader to (Arcidiacono and Miller, 2011), as well as our conclusion.

[6] Assume there are two clusters with a prior known cluster affiliation, then

$$ln\frac{Pr(z_{bt}^1)}{Pr(z_{bt}^2)} = \gamma_0^1 + \gamma_1^1 \mathbb{I}\left[s = 1\right],$$

with the probabilities almost equal to unity

$$\hat{\gamma}_1^1 = \infty \text{ and } Pr(z_b^1|\mathbb{I}\left[s = 1\right]; \hat{\boldsymbol{\gamma}}) \approx 1.$$

3.3.1 Observed likelihood

We parameterize equation 9 assuming productivity follows a Gaussian mixture. This specification is in line with Dewitte et al. (2020), who found the firm-size distribution to be best represented by a finite mixture of Log-normals.[7] The probability of observing a single observation is then

$$p^o(\phi_{bt}|k_{bt}, l_{bt}, \phi_{bt-1}, l_{bt-1}, k_{bt-1}, z_b^s; \boldsymbol{\beta}, \boldsymbol{\alpha}^s, \sigma_\eta^s) = \frac{1}{\sigma_\eta^s}\varphi\left(\frac{f^{kl}\left(k_{bt}, l_{bt}; \boldsymbol{\beta}\right) + g(\phi_{bt-1}, l_{bt-1}, k_{bt-1}; \boldsymbol{\beta}, \boldsymbol{\alpha^s})}{\sigma_\eta^s}\right). \tag{18}$$

It is possible that labor input is a dynamic, but not predetermined input (Ackerberg et al., 2015). In that case, it is common to instrument labor with its lagged value, which need to be taken into account when specifying the observed likelihood. The reduced-form specification for endogenous variable with exogenous instruments $k_{bt}, l_{bt-1}, \phi_{bt-1}$ and normally distributed error term ($\zeta_{bt} \sim \mathcal{N}\left(0, \sigma_\zeta\right)$) is specified as:

$$l_{bt} = \delta_0 + \delta_1 k_{bt} + \delta_2^s \phi_{bt-1} + \delta_3 k_{bt-1} + \delta_4^s l_{bt-1} + \zeta_{bt}. \tag{19}$$

Notice that we specify the capital-coefficients of the reduced-form instrumental equation to be cluster-independent, in line with the production function specification. Conditional on this reduced-form instrumental equation, the observed likelihood becomes

$$\begin{aligned} &p^o(\phi_{bt}|k_{bt}, l_{bt}, \phi_{bt-1}, l_{bt-1}, k_{bt-1}, z_b^s; \underbrace{\boldsymbol{\beta}, \boldsymbol{\alpha}^s, \sigma_\eta^s, \sigma_\zeta, \rho^s}_{=\boldsymbol{\theta}^s}) = \\ &\frac{1}{\sqrt{(1-(\rho^s)^2)}\sigma_\eta^s}\varphi\left(\frac{f^{kl}\left(k_{bt}, l_{bt}; \boldsymbol{\beta}\right) + g(\phi_{bt-1}, l_{bt-1}, k_{bt-1}; \boldsymbol{\beta}, \boldsymbol{\alpha^s}) + \rho\frac{\sigma_\eta^s}{\sigma_\zeta} l_{bt}\left(k_{bt}, \phi_{bt-1}, k_{bt-1}, l_{bt-1}; \boldsymbol{\delta}^s\right)}{\sqrt{(1-(\rho^s)^2)}\sigma_\eta^s}\right). \end{aligned} \tag{20}$$

where ρ^s represent the cluster-specific correlation coefficient between non-flexible output (ϕ_{bt}) and labor l_{bt}.[8]

Provided that our data consists of B time series, we can write the probability of observing such a time series $\boldsymbol{\phi} = \{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_B\} = \{\phi_{11}, \ldots, \phi_{1T}, \ldots, \phi_{B1}, \ldots, \phi_{BT}\}$ as[9]

[7] "A cluster of firms suggests a degree of *homogeneity* within a cluster and a degree of *separation* between clusters, while typically *not* implying any specific within-cluster distribution." (Fruhwirth-Schnatter et al., 2019, p.17) The number of components in a mixture model does not have to coincide with the number of clusters if the distributional assumption is incorrect.

[8] The corresponding joint distribution can be modeled as a Multivariate normal specification

$$\begin{bmatrix}\phi_{bt}\\ l_{bt}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} f^{kl}\left(k_{bt}, l_{bt}; \boldsymbol{\beta}\right) + g(\phi_{bt-1}, l_{bt-1}, k_{bt-1}; \boldsymbol{\beta}, \boldsymbol{\alpha^s}) \\ \delta_0 + \delta_1 k_{bt} + \delta_2^s \phi_{bt-1} + \delta_3 k_{bt-1} + \delta_4^s l_{bt-1}\end{bmatrix}, \begin{bmatrix}\sigma_\eta^2 & \rho^s \sigma_\eta \sigma_\zeta \\ \rho^s \sigma_\eta \sigma_\zeta & \sigma_\zeta^2\end{bmatrix}\right)$$

[9] Following the literature, we discard of the first time period data and specifying the initial condition.

$$p^o(\boldsymbol{\phi}_b|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b, z_b^s; \boldsymbol{\theta}^s) = \prod_{t=1}^{T} p(\phi_{bt}|k_{bt}, l_{bt}, \phi_{bt-1}, l_{bt-1}, k_{bt-1}, z_b^s; \boldsymbol{\theta}^s). \tag{21}$$

Subsequently, the probability of observing the complete data series can be defined as

$$p^o(\boldsymbol{\phi}; \underbrace{\{\boldsymbol{\gamma}^1, \ldots, \boldsymbol{\gamma}^S, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^S\}}_{=\boldsymbol{\Theta}}) = \prod_{b=1}^{B} \sum_{s=1}^{S} Pr(z_b^s|k_{b0}, l_{b0}, \omega_{b0}; \boldsymbol{\gamma}^s) p^o(\boldsymbol{\phi}_b|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b, z_b^s; \boldsymbol{\theta}^s). \tag{22}$$

$$\tag{23}$$

3.3.2 Complete log-likelihood

From equation 18, we can specify the observed log-likelihood

$$\mathcal{L}^o(\boldsymbol{\Theta}) = \sum_{b=1}^{B} \sum_{s=1}^{S} log\left(Pr(z_b^s|k_{b0}, l_{b0}, \omega_{b0}; \boldsymbol{\gamma}^s) p^o(\boldsymbol{\phi}_b|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b, z_b^s; \boldsymbol{\theta}^s)\right), \tag{24}$$

which does not account for the unobserved cluster affiliation. Additionally accounting for cluster affiliation (z_b^s), we arrive at the complete log-likelihood

$$\mathcal{L}^c(\boldsymbol{\Theta}, \boldsymbol{z}) = \sum_{b=1}^{B} \sum_{s=1}^{S} z_b^s log\left(Pr(z_b^s|k_{b0}, l_{b0}, \omega_{b0}; \boldsymbol{\gamma}^s) p^o(\boldsymbol{\phi}_b|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b, z_b^s; \boldsymbol{\theta}^s)\right), \tag{25}$$

which forms the basis for our estimation procedure.

3.3.3 Estimation procedure

We estimate the parameters of interest based on equation 21 relying on the Expectation-Maximization algorithm (McLachlan and Peel, 2000; Miljkovic and Grün, 2016). This algorithm consists of maximizing the complete log-likelihood in an iterative procedure. Assume parameter values in iteration j are represented as $(\boldsymbol{\Theta})^j = \left\{(\boldsymbol{\gamma}^1)^j, \ldots, (\boldsymbol{\gamma}^S)^j, (\boldsymbol{\theta}^1)^j, \ldots, (\boldsymbol{\theta}^S)^j\right\}$.

1. In a first step, we approximate cluster affiliation with the posterior conditional probability obtained from Bayes' theorem:

$$\hat{z}_b^s = Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; (\boldsymbol{\Theta})^j) = \frac{Pr(z_b^s|k_{b0}, l_{b0}, \omega_{b0}; (\boldsymbol{\gamma}^s)^j) p^o(\boldsymbol{\phi}_b|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b, z_b^s; (\boldsymbol{\theta}^s)^j)}{p^o(\boldsymbol{\phi}; (\boldsymbol{\Theta})^j)} \tag{26}$$

2. In a second step, these approximations of cluster affiliation are relied upon for to estimate the parameters $(\boldsymbol{\Theta})^{j+1}$:

   –

$$\max_{(\boldsymbol{\theta})^{j+1}} \sum_{b=1}^{B} \sum_{s=1}^{S} Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; (\boldsymbol{\Theta})^j) log\left(p^o(\boldsymbol{\phi}_b|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b, z_b^s; (\boldsymbol{\theta}^s)^{j+1})\right)$$

–

$$\max_{(\boldsymbol{\theta})^{j+1}} \sum_{b=1}^{B}\sum_{s=1}^{S} Pr(z_b^s|\boldsymbol{k}_b,\boldsymbol{l}_b,\boldsymbol{\phi}_b;(\boldsymbol{\Theta})^j) log\left(p^o(\boldsymbol{\phi}_b|\boldsymbol{k}_b,\boldsymbol{l}_b,\boldsymbol{\phi}_b,z_b^s;(\boldsymbol{\theta}^s)^{j+1})\right)$$

–

$$\max_{(\boldsymbol{\gamma}^s)^{j+1}} \sum_{b=1}^{B}\sum_{s=1}^{S} Pr(z_b^s|\widehat{\boldsymbol{k}_b,\boldsymbol{l}_b,\boldsymbol{\phi}}_b;(\boldsymbol{\Theta})^j) log\left(Pr(z_b^s|k_{b0},l_{b0},\omega_{b0};(\boldsymbol{\gamma}^s)^{j+1})\right)$$

This iterative process is continued untill there is relative stability between iteration j and $j+1$ in terms of the observed log-likelihood (eq. 20).

3.4 Monte Carlo

We rely on a Monte Carlo (MC) exercise to evaluate the performance of the proposed estimator. We focus on the estimators' ability to recover unobserved heterogeneity in the productivity distribution while affirming the importance of controlling for unobserved heterogeneity in production function estimations. The setup of the Monte Carlo exercise closely mimics the setup of Ackerberg et al. (2015) that builds on (Syverson, 2001; Van Biesebroeck, 2007). It only differs from Ackerberg et al. (2015) in the specification of the Markov process of productivity, which is assumed to differ between clusters of firms, as is described below.

3.4.1 Model Setup

The data used for this project is a simulated panel dataset of 1,000 firms over 10 years.

Production function and productivity shocks This data is constructed assuming a Leontief production function:

$$Y_{it} = min\left\{K_{it}^{\beta_k}L_{it}^{\beta_l}e^{\omega_{it}}, \beta_m M_{it}\right\}e^{\epsilon_{it}} \tag{27}$$

where $\beta_k = 0.4$, $\beta_l = 0.6$, and $\beta_m = 1$, implying proportionality between output Y_{it} and material input M_{it}. ϵ_{it} is measurement error that is normally distributed, $\mathcal{N}(0, 0.1)$. In contrast to Ackerberg et al. (2015), Log-productivity, ω_{it}, follows an *Finite Mixture* AR(1)-process

$$\omega_{it} = \sum_{c=1}^{2}\pi^c\left[\alpha^s + \alpha^s\omega_{it-1} + \eta_{it}^s\right],$$

with 800 observations assigned to $c = 1$, $\pi^1 = 0.8$, and 200 observations to $c = 2$, $\pi^2 = 0.2$. We assume $\eta_{it} \sim \mathcal{N}\left(0, \sigma_\eta^k\right)$.

Choice of Labor and Material inputs We follow the first Data Generating Process (DGP) of (Ackerberg et al., 2015) for the labor (and material) inputs. Labor and materials are assumed to be flexible inputs (see Section ...). There are firm-specific (unobserved to the econometrician) wage shocks, and L_{it} is chosen prior to period t, that is, without full knowledge of ω_{it}. More specifically, labor is chosen at time period $t - b$ (we set $b = 0.5$). We can think of decomposing the Finite Mixture AR(1)-process (eq. 3.4.1) into two subprocesses. First, ω_{it-1} evolves to $\omega_{i,t-b}$, at which point in time the firm chooses labor input (as a function of $\omega_{i,t-b}$). Then, after L_{it} is chosen, $\omega_{i,t-b}$ evolves to ω_{it}.

The evolution of ω between subperiod is specified as follows:

$$\begin{aligned}\omega_{i,t-b} &= \sum_{c=1}^{2} \pi^c \left[\alpha_0^c + \left(\alpha_1^c\right)^{1-b} \omega_{i,t-1} + \eta_{it}^{c,A}\right] \\ \omega_{it} &= \sum_{c=1}^{2} \pi^c \left[\left(1-\alpha_1^b\right)\alpha_0^c + \left(\alpha_1^c\right)^{b} \omega_{i,t-b} + \eta_{it}^{c,B}\right] \end{aligned} \tag{28}$$

Thus, when $b > 0$, firms have less than perfect information about ω_{it} when choosing L_{it}, and when b increases, this information decreases. Note that this specification is consistent with the Finite Mixture AR(1)-process specified in equation 3.4.1 since $\left(1-\alpha_1^b\right)\alpha_0^c + \alpha_1^b\alpha_0^c = \alpha_0^c$ and $\left(\alpha_1^c\right)^{1-b}\left(\alpha_1^c\right)^{b} = \alpha_1^c$. Additionally, we follow Ackerberg et al. (2015) in imposing that $Var\left(\left(\alpha_1^c\right)^b \eta_{it}^{c,A} + \eta_{it}^{c,B}\right) = Var\left(\eta_{it}^c\right)$ and that the variance of $\eta_{it}^{c,A}$ is such that the variance of $\omega_{i,t-b}$ is constant over time. This defines $Var\left(\eta_{it}^{c,A}\right) = \sigma^2_{\eta^{c,A}}$ and $Var\left(\eta_{it}^{c,B}\right) = \sigma^2_{\eta^{c,B}}$.

Firms also face different wages where the log-wage process for firm i follows an AR(1)-process

$$\ln\left(W_{it}\right) = 0.3\ln\left(W_{it-1}\right) + \eta_{it}^W \tag{29}$$

where the variances of the normally distributed innovation $\eta_{it}^W\left(\sigma^2_{\eta^W}\right)$ and the initial value $\ln\left(W_{i0}\right)$ are set such that the standard deviation of $\ln\left(W_{it}\right)$ is constant over time and equal to 0.1. Relative to a baseline in which all firms face the mean log wage in every period, this wage variation increases the within-firm, across-time, standard deviation of $\ln\left(L_{it}\right)$ by about 10% (Ackerberg et al., 2015).

Given this DGP, firms optimally choose L_{it} to maximize expected profits by setting (with the difference between the price of output and the price of the material input normalized to 1)

$$L_{it} = \beta_l^{1/(1-\beta_l)} W_{it}^{-1/(1-\beta_l)} K_{ii}^{\beta_k/(1-\beta_l)} e^{(1/(1-\beta_l))\left(\alpha_0^c + \left(\alpha_1^c\right)^b \omega_{it-1} + (1/2)\sigma^2_{\eta^{c,B}}\right)}$$

Investment choice and steady state In contrast to the flexible labor and material inputs, capital is assumed to be a dynamic input. Specifically, capital is accumulated through investment according to

$$K_{it} = (1-\delta)K_{it-1} + I_{it-1}, \tag{30}$$

where $(1-\delta) = 0.8$, and investment is subject to convex adjustment costs given by

$$c_i\left(I_{it}\right)=\frac{\phi_i}{2}I_{it}^2 \tag{31}$$

where $1/\phi_i$ is distributed lognormally across firms (but constant over time) with standard deviation 0.6.

Under the assumption of constant returns to scale, a pared-down version of the above can be solved analytically using Euler equation techniques. Specifically, a Euler equation approach implies the following optimal investment rule (where β is the discount factor, set to 0.95 in the Monte Carlos):

$$\begin{aligned}
I_{it}=&\frac{\beta}{\phi_i}\sum_{\tau=0}^{\infty}(\beta(1-\delta))^{\tau}\left(\frac{\beta_k}{1-\beta_l}\right)\\
&\times\left[\beta_l^{\beta_l/(1-\beta_l)}e^{(1/2)\beta_l^2\sigma_l^2}-\beta_l^{1/(1-\beta_l)}e^{(1/2)\sigma_l^2}\right]\\
&\times\exp\left\{\left[\left(\frac{1}{1-\beta_l}\right)\alpha_0^c+\left(\frac{1}{1-\beta_l}\right)(\alpha_1^c)^{\tau+1}\omega_{it}+\frac{-\beta_l}{1-\beta_l}\rho_W^{\tau+1}\ln\left(W_{it}\right)\right.\right.\\
&+\frac{1}{2}\left(\frac{-\beta_l}{1-\beta_l}\right)^2\sigma_{\xi^W}^2\sum_{s=0}^{\tau}\rho_W^{2(\tau-s)}+\frac{1}{2}\left(\frac{1}{1-\beta_l}\right)^2(\alpha_1^c)^{2b}\left((\alpha_1^c)^{2\tau}\sigma_{\eta^{c,A}}^2\right.\\
&\left.\left.\left.+\sum_{s=1}^{\tau}\rho^{2(\tau-s)}\sigma_\eta^2\right)+\left(\frac{1}{1-\beta_l}\right)\left(\frac{1}{2}\sigma_{\eta^{c,B}}^2\right)\right]\right\}
\end{aligned} \tag{32}$$

To avoid dependence on the initial conditions, the data is simulated over one hundred periods of which only the last 10 periods of data are withheld.

3.4.2 Data Generating process

We specify three data generating processes.

- **DGP1 - Mixture, Endogeneity**: $c^1=1,\ c^2=1.2$ and $b=0.5$ and $Pr(z_b^s|k_{b0},l_{b0},\omega_{b0})$;
- **DGP2 - Proxy variables**: $c^1=1,\ c^2=1.2$ and $b=0.5$ and $Pr(z_b^s|k_{b0},l_{b0},\omega_{b0},\boldsymbol{e}_b^s)$;
- **DGP3 - Noisy proxy variables**: $c^1=1,\ c^2=1.2$, $b=0.5$ and $Pr(z_b^s|k_{b0},l_{b0},\omega_{b0},\boldsymbol{e}_b^s)$ with 10% of the proxy variable misclassified.

3.4.3 Results

Table 2: Monte Carlo results - DGP1

Methodology	β_k	β_l	c_1	ρ_1	σ_1	c_2	ρ_2	σ_2	π_1	π_2
True coefficients	0.40	0.60	1.00	0.70	0.21	1.20	0.70	0.21	0.80	0.20
	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
NLLS	-0.02	1.05	1.01	0.05	0.17	-	-	-	1.00	1.00
	(0.01)	(0.00)	(0.02)	(0.01)	(0.00)	(-)	(-)	(-)	(0.00)	(0.00)
ML	-0.02	1.05	1.01	0.05	0.17	-	-	-	1.00	1.00
	(0.01)	(0.00)	(0.02)	(0.01)	(0.00)	(-)	(-)	(-)	(0.00)	(0.00)
GMM	0.54	0.58	0.48	0.82	0.22	-	-	-	1.00	1.00
	(0.03)	(0.01)	(0.02)	(0.01)	(0.00)	(-)	(-)	(-)	(0.00)	(0.00)
LIML	0.50	0.58	0.50	0.82	0.22	-	-	-	1.00	1.00
	(0.02)	(0.01)	(0.02)	(0.01)	(0.00)	(-)	(-)	(-)	(0.00)	(0.00)
Mixture NLLS	-0.00	1.00	1.51	-0.00	0.18	1.73	-0.01	0.18	0.80	0.20
	(0.01)	(0.00)	(0.03)	(0.01)	(0.00)	(0.09)	(0.05)	(0.00)	(0.00)	(0.00)
Mixture ML	-0.00	1.00	1.51	-0.00	0.18	1.72	-0.00	0.18	0.80	0.20
	(0.01)	(0.00)	(0.03)	(0.01)	(0.00)	(0.04)	(0.02)	(0.00)	(0.00)	(0.00)
Mixture GMM	0.39	0.60	1.03	0.69	0.22	1.26	0.69	0.22	0.80	0.20
	(0.02)	(0.01)	(0.04)	(0.01)	(0.00)	(0.18)	(0.04)	(0.01)	(0.01)	(0.01)
Mixture LIML	0.39	0.60	1.03	0.69	0.22	1.15	0.71	0.22	0.79	0.21
	(0.02)	(0.01)	(0.04)	(0.01)	(0.00)	(0.06)	(0.02)	(0.00)	(0.01)	(0.01)

Table 3: Monte Carlo results - DGP2

Methodology	β_k	β_l	c_1	ρ_1	σ_1	c_2	ρ_2	σ_2	π_1	π_2
True coefficients	0.40	0.60	1.00	0.70	0.21	1.20	0.70	0.21	0.80	0.20
	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
NLLS	-0.02	1.05	1.01	0.05	0.17	-	-	-	1.00	1.00
	(0.01)	(0.00)	(0.02)	(0.01)	(0.00)	(-)	(-)	(-)	(0.00)	(0.00)
ML	-0.02	1.05	1.01	0.05	0.17	-	-	-	1.00	1.00
	(0.01)	(0.00)	(0.02)	(0.01)	(0.00)	(-)	(-)	(-)	(0.00)	(0.00)
GMM	0.54	0.58	0.48	0.82	0.22	-	-	-	1.00	1.00
	(0.03)	(0.01)	(0.02)	(0.01)	(0.00)	(-)	(-)	(-)	(0.00)	(0.00)
LIML	0.50	0.58	0.50	0.82	0.22	-	-	-	1.00	1.00
	(0.02)	(0.01)	(0.02)	(0.01)	(0.00)	(-)	(-)	(-)	(0.00)	(0.00)
Mixture NLLS	-0.00	1.00	1.50	-0.00	0.18	1.73	-0.01	0.18	0.80	0.20
	(0.01)	(0.00)	(0.03)	(0.01)	(0.00)	(0.09)	(0.05)	(0.00)	(0.00)	(0.00)
Mixture ML	-0.00	1.00	1.50	-0.00	0.18	1.70	-0.00	0.18	0.80	0.20
	(0.01)	(0.00)	(0.03)	(0.01)	(0.00)	(0.05)	(0.02)	(0.00)	(0.00)	(0.00)
Mixture GMM	0.40	0.60	1.01	0.70	0.22	1.25	0.69	0.22	0.80	0.20
	(0.02)	(0.01)	(0.03)	(0.01)	(0.00)	(0.17)	(0.04)	(0.00)	(0.00)	(0.00)
Mixture LIML	0.40	0.60	1.02	0.69	0.22	1.22	0.70	0.22	0.80	0.20
	(0.02)	(0.01)	(0.03)	(0.01)	(0.00)	(0.05)	(0.02)	(0.00)	(0.00)	(0.00)

Table 4: Monte Carlo results - DGP3

Methodology	β_k	β_l	c_1	ρ_1	σ_1	c_2	ρ_2	σ_2	π_1	π_2
True coefficients	0.40	0.60	1.00	0.70	0.21	1.20	0.70	0.21	0.80	0.20
	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
NLLS	-0.02	1.05	1.00	0.05	0.17	-	-	-	1.00	1.00
	(0.01)	(0.00)	(0.03)	(0.01)	(0.00)	(-)	(-)	(-)	(0.00)	(0.00)
ML	-0.02	1.05	1.01	0.05	0.17	-	-	-	1.00	1.00
	(0.01)	(0.00)	(0.03)	(0.01)	(0.00)	(-)	(-)	(-)	(0.00)	(0.00)
GMM	0.53	0.57	0.48	0.82	0.22	-	-	-	1.00	1.00
	(0.02)	(0.01)	(0.02)	(0.01)	(0.00)	(-)	(-)	(-)	(0.00)	(0.00)
LIML	0.50	0.58	0.50	0.83	0.22	-	-	-	1.00	1.00
	(0.02)	(0.01)	(0.02)	(0.01)	(0.00)	(-)	(-)	(-)	(0.00)	(0.00)
Mixture NLLS	0.00	1.00	1.50	-0.00	0.18	1.69	0.01	0.18	0.80	0.20
	(0.01)	(0.00)	(0.04)	(0.01)	(0.00)	(0.10)	(0.05)	(0.00)	(0.00)	(0.00)
Mixture ML	-0.00	1.00	1.50	-0.00	0.18	1.70	0.00	0.18	0.80	0.20
	(0.01)	(0.00)	(0.04)	(0.01)	(0.00)	(0.06)	(0.02)	(0.00)	(0.00)	(0.00)
Mixture GMM	0.40	0.60	1.01	0.70	0.22	1.25	0.69	0.22	0.80	0.20
	(0.01)	(0.01)	(0.04)	(0.01)	(0.00)	(0.21)	(0.05)	(0.01)	(0.00)	(0.00)
Mixture LIML	0.40	0.60	1.01	0.70	0.22	1.17	0.71	0.22	0.80	0.20
	(0.01)	(0.01)	(0.03)	(0.01)	(0.00)	(0.07)	(0.02)	(0.00)	(0.00)	(0.00)

4 Application to Belgian firm-level data

4.1 Data

This methodology set out in this paper is applied to a database of Belgian manufacturing firms over the period 2008-2018 that gathers a very general set of firm-level characteristics considered relevant for productivity growth.[10] This includes, but is not limited to: firm age, firm sector affiliation, firm location, firm-level exporting and importing intensity, indicators of management quality, whether or not the firm is a single- or multi-establishment firm, number of products produced by the firm, the emission caps and allowances of the firm, whether or not the firm is part of or participates in FDI, the intensity of upstream- and downstream product flows, the amount of Research and Development undertaken by the firm, and financial constraints of the firm. A considerable effort has been made to construct this extensive database relying on resources from, among others, the NBB, Orbis Europe, the Belgian Science Policy Office, and the European Patent Office.

4.2 Results

We estimate in a first stage

$$va_{bt} = f(k_{bt}, l_{bt}, m_{bt}) + \epsilon_{bt} \tag{33}$$

so that we can identify the parameters of interest from a second stage

$$\phi_{bt} = va_{bt} - \epsilon_{bt} = \beta_k k_{bt} + \beta_l l_{bt} + \sum_{s=1}^{S} \pi^s \left[\alpha_0^s + \alpha_1^s \omega_{bt-1} + \eta_{bt}^s\right]. \tag{34}$$

We rely both on the traditional NLLS and GMM without additional heterogeneity in the Markov process, as well as on LIML with increasing heterogeneity (nr. clusters) in the Markov process.

[10] A similar but less extensive database has already been used for productivity estimations by, among others, (Forlani et al., 2016; De Loecker et al., 2014; Mion and Zhu, 2013).

Table 5: ACF estimation results

Method	β_k	β_l	π^s	α_0^s	α_1^s	σ_η^s	**logl**	**AIC**	**BIC**
NLLS	0.13	0.80	1.00	0.52	0.95	0.09	108,615.9	-217,221.9	-217,174.0
GMM	0.10	0.90	1.00	0.64	0.94	0.10	100,070.9	-200,131.8	-200,083.8
LIML 1	0.09	0.92	1.00	0.59	0.94	0.10	126,043.8	-252,077.5	-252,029.6
LIML 2	0.08	0.94	0.60	0.43	0.96	0.06	167,505.5	-334,982.9	-334,848.7
			0.40	0.77	0.92	0.15			
LIML 3	0.08	0.95	0.44	0.42	0.96	0.06	177,629.8	-355,213.7	-354,993.2
			0.35	0.48	0.95	0.09			
			0.21	1.08	0.89	0.19			
LIML 4	0.08	0.96	0.33	0.49	0.95	0.10	182,917.1	-365,770.2	-365,463.4
			0.32	0.40	0.96	0.04			
			0.21	0.66	0.93	0.10			
			0.14	1.21	0.88	0.21			
LIML 5	0.08	0.97	0.25	0.52	0.95	0.12	185,701.9	-371,321.8	-370,928.8
			0.23	0.38	0.96	0.04			
			0.23	0.44	0.96	0.06			
			0.20	0.68	0.93	0.10			
			0.09	1.46	0.85	0.24			
LIML 6	0.10	0.96	0.26	0.46	0.95	0.07	187,105.9	-374,111.8	-373,632.5
			0.22	0.55	0.94	0.13			
			0.19	0.36	0.96	0.04			
			0.18	0.52	0.95	0.07			
			0.10	0.82	0.91	0.16			
			0.06	0.93	0.90	0.26			

5 Conclusion

...

References

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification Properties of Recent Production Function Estimators. *Econometrica 83*(6), 2411–2451. 1, 2, 4, 6, 8, 9, 11, 12

Amiti, M. and J. Konings (2007). Trade liberalization, intermediate inputs, and productivity: Evidence from indonesia. *American Economic Review 97*(5), 1611–1638. 3

Amsler, C., A. Prokhorov, and P. Schmidt (2016). Endogeneity in stochastic frontier models. *Journal of Econometrics 190*(2), 280–288. 4

Arcidiacono, P. and J. B. Jones (2003). Finite Mixture Distributions, Sequential Likelihood and the EM Algorithm. *Econometrica 71*(3), 933–946. 6

Arcidiacono, P. and R. A. Miller (2011). Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica 79*(6), 1823–1867. 8

Atkeson, A. and A. Burstein (2010). Innovation, firm dynamics, and international trade. *Journal of Political Economy 118*(3), 433–484. 1

Aw, B. Y., M. J. Roberts, and D. Y. Xu (2011). R&d investment, exporting, and productivity dynamics. *American Economic Review 101*(4), 1312–44. 3

Battisti, M., F. Belloc, and M. Del Gatto (2020). Is the productivity premium of internationalized firms technology-driven? *Empirical Economics*, 1–34. 4

Battisti, M., F. Belloc, and M. Del Gatto (2020). Labor productivity and firm-level tfp with technology-specific production functions. *Review of Economic Dynamics 35*, 283–300. 4

Beard, T. R., S. B. Caudill, and D. M. Gropper (1997). The diffusion of production processes in the us banking industry: A finite mixture approach. *Journal of Banking & Finance 21*(5), 721–740. 4

Bee, M., M. Riccaboni, and S. Schiavo (2011, Aug). Pareto versus lognormal: A maximum entropy test. *Phys. Rev. E 84*, 026104. 1, 3

Blalock, G. and P. J. Gertler (2008). Welfare gains from foreign direct investment through technology transfer to local suppliers. *Journal of International Economics 74*(2), 402–421. 3

Bloom, N. and J. Van Reenen (2011). Human resource management and productivity. In *Handbook of labor economics*, Volume 4, pp. 1697–1767. Elsevier. 1, 3

Caliendo, L., G. Mion, L. D. Opromolla, and E. Rossi-Hansberg (2015). Productivity and organization in portuguese firms. Working Paper 21811, National Bureau of Economic Research. 1, 3

Celeux, G. and G. Govaert (1992). A classification em algorithm for clustering and two stochastic versions. *Computational Statistics & Data Analysis 14*(3), 315–332.

Costantini, J. and M. Melitz (2008). The dynamics of firm-level adjustment to trade liberalization. *The organization of firms in a global economy 4*, 107–141. 1

Das, S., M. J. Roberts, and J. R. Tybout (2007). Market entry costs, producer heterogeneity, and export dynamics. *Econometrica 75*(3), 837–873. 3

De Loecker, J. (2013, August). Detecting learning by exporting. *American Economic Journal: Microeconomics 5*(3), 1–21. 1, 3, 7

De Loecker, J., C. Fuss, and J. Van Biesebroeck (2014). International competition and firm performance: Evidence from belgium. Working Paper 269, National Bank of Belgium. 16

Dewitte, R., M. Dumont, G. Rayp, and P. Willemé (2020). Unobserved heterogeneity in the productivity distribution and gains from trade. MPRA Paper 102711, Ghent University. 9

Doraszelski, U. and J. Jaumandreu (2013). R&d and productivity: Estimating endogenous productivity. *The Review of Economic Studies 80*(4), 1338–1383. 3

El-Gamal, M. A. and H. Inanoglu (2005). Inefficiency and heterogeneity in turkish banking: 1990–2000. *Journal of Applied Econometrics 20*(5), 641–664. 4

European Commission (2015). The 2015 ageing report. economic and budgetary projections for the 28 eu member states (2013–2060). European Economy series 3, Directorate-General for Economic and Financial Affairs. 1

Follmann, D. A. and D. Lambert (1991). Identifiability of finite mixtures of logistic regression models. *Journal of Statistical Planning and Inference 27*(3), 375–381.

Forlani, E., R. Martin, G. Mion, and M. Muuls (2016). Unraveling firms: Demand, productivity and markups heterogeneity. Working Paper 293, National Bank of Belgium. 1, 2, 16

Fruhwirth-Schnatter, S., G. Celeux, and C. P. Robert (2019). *Handbook of mixture analysis*. CRC press. 9

Gandhi, A., S. Navarro, and D. Rivers (2017). How Heterogeneous is Productivity? A Comparison of Gross Output and Value Added? Working paper. 5

Gandhi, A., S. Navarro, and D. A. Rivers (2020). On the identification of gross output production functions. *Journal of Political Economy 128*(8), 2973–3016. 1, 2, 4, 6

Gormley, I. C. and S. Frühwirth-Schnatter (2019). Mixture of experts models. *Handbook of mixture analysis*, 271–307. 4

Greene, W. (2005). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics 126*(2), 269–303. 4

Grün, B. (2018). Model-based clustering. arXiv preprint 1807.01987, arXiv.

Grün, B. and F. Leisch (2008). Identifiability of finite mixtures of multinomial logit models with varying and fixed effects. *Journal of classification 25*(2), 225–247.

Hennig, C. (2000). Identifiablity of models for clusterwise linear regression. *Journal of classification 17*(2).

Javorcik, B. S. (2004). Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers through Backward Linkages. *The American Economic Review 94*(3), 605–627. 3

Kasahara, H. and B. Lapham (2013). Productivity and the decision to import and export: Theory and evidence. *Journal of International Economics 89*(2), 297–316. 1, 3

Kasahara, H. and J. Rodrigue (2008). Does the use of imported intermediates increase productivity? plant-level evidence. *Journal of Development Economics 87*(1), 106–118. 1, 3

Kasahara, H., P. Schrimpf, and M. Suzuki (2017). Identification and estimation of production function with unobserved heterogeneity. *mimeo*. 3, 4

Levinsohn, J. and A. Petrin (2003). Estimating productin functions using inputs to control for unobservables. *The Review of Economic Studies 70*, 317–341. 1, 2

Luttmer, E. G. (2007). Selection, growth, and the size distribution of firms. *The Quarterly Journal of Economics 122*(3), 1103–1144. 1

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, 105–142. 8

McLachlan, G. J. and D. Peel (2000). *Finite mixture models*. New York: Wiley Series in Probability and Statistics. 10

Miljkovic, T. and B. Grün (2016). Modeling loss data using mixtures of distributions. *Insurance: Mathematics and Economics 70*, 387 – 396. 10

Mion, G. and L. Zhu (2013). Import competition from and offshoring to china: A curse or blessing for firms? *Journal of International Economics 89*(1), 202–215. 16

Olley, S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica 64*(6), 1263–1297. 1, 2, 3

Orea, L. and S. C. Kumbhakar (2004). Efficiency measurement using a latent class stochastic frontier model. *Empirical Economics 29*(1), 169–183. 4

Peersman, G. (2019, November). Financiering en betaalbaarheid van de sociale zekerheid. Commissie voor sociale zaken, werk en pensioenen. 1

Sickles, R. C. and V. Zelenyuk (2019). *Measurement of Productivity and Efficiency*. Cambridge University Press. 4

Syverson, C. (2001). *Market structure and productivity*. Ph.d. dissertation, University of Maryland. 11

Van Beveren, I. (2012). Total factor productivity estimation: A practical review. *Journal of economic surveys 26*(1), 98–128. 2

Van Beveren, I. and S. Vanormelingen (2014). Human capital, firm capabilities and productivity growth. Working Paper 257, National Bank of Belgium. 3

van Biesebroeck, J. (2003). Productivity Dynamics with Technology Choice: An Application to Automobile Assembly. *The Review of Economic Studies 70*(1), 167–198. 3, 4

Van Biesebroeck, J. (2007). Robustness of productivity estimates. *The Journal of Industrial Economics 55*(3), 529–569. 11

Online Appendix to "Identifying Unobserved Heterogeneity in Productivity"

Ruben Dewitte, Catherine Fuss and Angelos Theodorakopoulos*

30th April 2021

*Dewitte (corresponding author): Ghent University, Tweekerkenstraat 2, 9000 Ghent, Belgium, rubenl.dewitte@ugent.be, Fuss: National Bank of Belgium, boulevard de Berlaimont 14, 1000 Brussels, Catherine.Fuss@nbb.be , Theodorakopoulos: Oxford Martin School, University of Oxford, 34 Broad Street, Oxford OX1 3BD, United Kingdom, angelos.theodorakopoulos@oxfordmartin.ox.ac.uk.

Appendix A Additional Figures and Table

A.1 Tables

Table 1: Data vs OECD structural SDBS database

	Number of Enterprises						Total Employment						Turnover					
Industry	1-9	10-19	20-49	50-249	> 250	Total	1-9	10-19	20-49	50-249	> 250	Total	1-9	10-19	20-49	50-249	> 250	Total
10	30	60	82	80	79	38	38	60	83	80	76	70	69	81	96	74	76	77
11	24	107	92	80	100	39	40	113	91	80	100	94	101	125	85	75	99	94
12			40			112			45			102			47			103
13	38	84	81	81	85	49	59	76	83	83	90	83	89	91	72	79	84	81
14	21	125	68			28	40	105	67			74	116	91	56			92
15	20	133				28	32	130				81	85	172				97
16	21	99	92	75	100	29	39	98	90	84	100	78	66	107	85	88	94	88
17	43	64	94	79	86	59	67	68	94	81	85	83	16	62	100	87	92	86
18	23	83	106	83	67	28	39	89	86	80	71	67	59	90	87	76	80	76
19					50	29					42	46					60	60
20	45	71	88	90	93	66	61	65	75	87	89	87	45	69	116	92	100	95
21	18	160	53	100	82	46	23	137	56	108	77	79	31	273	69	101	87	87
22	45	105	93	86	82	65	71	98	88	84	85	85	81	109	83	84	68	79
23	41	76	88	82	96	52	59	74	92	77	87	81	78	98	104	86	97	93
24		80	96		87	40		78	95		96	93		54	405		92	91
25	28	77	85	80	71	37	51	80	81	78	78	74	68	77	84	79	72	77
26	17	110	76	92	100	33	30	115	76	93	100	91	44	123	53	93	99	90
27	51	79	96	93	56	64	77	80	92	91	52	67	128	71	106	100	66	78
28	37	82	89	83	80	52	55	80	86	85	81	81	76	77	89	101	68	79
29	34	70	124	86	96	51	48	78	100	92	76	78	56	102	78	76	64	66
30	24	143	73	75	100	40	42	139	71	81	100	95	64	243	91	79	105	103
31	26	82	81	80	100	34	49	85	81	81	100	75	63	91	98	75	103	82
32	20	102	110	78	67	24	36	95	91	64	82	63	70	106	98	34	108	78
33	19	66	83	87	118	27	37	63	81	84	148	99	47	50	81	71	119	88
All	28	77	88	87	87	37	45	77	84	87	85	79	45	77	84	87	85	79

Each cell corresponds to the ratio of our dataset compared to the data from the OECD structural SDBS database for the year 2013. Size classes are based on total employment in FTE equivalents.