Identifying Latent Heterogeneity in Productivity

Ruben Dewitte, Catherine Fuss and Angelos Theodorakopoulos*

December, 2022

Abstract

Productivity is influenced by several firm-level factors, often latent. When unexplained, this latent heterogeneity can lead to the mismeasurement of productivity differences between groups of firms. We propose a flexible, semi-parametric extension of current production function estimation techniques using finite mixture models to control for latent firm-specific productivity determinants. We establish the performance of the proposed methodology through a Monte Carlo analysis and estimate export premia using firm-level data to demonstrate its empirical applicability. We apply our framework to assess export productivity premia and their robustness with respect to latent heterogeneity. Our results highlight that latent heterogeneity distorts export premia estimates and their contribution to aggregate productivity growth. The proposed approach delivers robust estimates of productivity differences between firm groups, regardless of the availability of productivity determinants in the data.

Keywords: finite mixture model, productivity estimation, productivity distribution, latent productivity determinants

JEL Codes: C13, C14, D24, L11

*Dewitte (corresponding author): Ghent University, Tweekerkenstraat 2, 9000 Ghent, Belgium, rubenl.dewitte@ugent.be; Fuss: National Bank of Belgium, 14 Boulevard de Berlaimont, 1000 Brussels, Belgium, catherine.fuss@nbb.be; Theodorakopoulos: Aston Business School, 295 Aston Express Way, Birmingham B4 7ER, UK, a.theodorakopoulos2@aston.ac.uk. The authors would like to thank Emmanuel Dhyne, Cédric Duprez, Rebecca Freeman, David Rivers, an anonymous referee and participants in the 2021 NAPW conference, the 2021 ITIM seminar at Ghent University, the 2022 Antwerp University seminar, the 2022 BORDERS workshop at Ghent University and the 2022 CEA conference at Carleton University. Ruben Dewitte gratefully acknowledges financial support from the Research Foundation Flanders (FWO) grant 12B8822N and the National Bank of Belgium (NBB). This paper was completed during the NBB's internship program for young researchers. The views expressed herein are those of the authors and do not necessarily reflect the views of the NBB or any other institution to which the authors are affiliated. Any errors are the authors' own.

1 Introduction

Firm-level productivity has been shown to differ between clusters of firms. These differences can be driven by a host of firm-level characteristics, such as innovation (Costantini and Melitz, 2008; Atkeson and Burstein, 2010; Bee et al., 2011), management practices (Bloom and Van Reenen, 2011; Caliendo et al., 2020), trade (Amiti and Konings, 2007; Kasahara and Rodrigue, 2008; De Loecker, 2013; Kasahara and Lapham, 2013) and industry linkages (Luttmer, 2007). Many characteristics remain latent and thus unobserved by the researcher. Therefore, productivity differences between groups of firms can be misidentified when latent heterogeneity in the productivity growth process is left unexplained (De Loecker, 2013).

This paper proposes a new methodology to estimate productivity, one that factors in heterogeneity in productivity growth originating from latent, time-invariant firm-level determinants. We build on the observation of Dewitte et al. (2022) that, in standard production function estimation methodologies, productivity for all firms is assumed to follow a homogeneous random growth process (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015; Gandhi et al., 2020). Therefore, if heterogeneity from firm-level characteristics is present but not controlled for, the level, dynamics, or drivers of productivity may be inaccurately identified. To circumvent these challenges, we propose an extension of current productivity estimation methodologies using finite mixture models (FMM). An FMM is a probabilistic model that allows for the productivity evolution to differ across clusters of firms, even when the immediate drivers of these differences are unobserved by researchers.

The proposed methodology builds on the behavioral framework set out by Olley and Pakes (1996) and adapted by Levinsohn and Petrin (2003); Doraszelski and Jaumandreu (2013); Ackerberg et al. (2015), and Gandhi et al. (2020), among others. Specifically, as commonly modeled in the literature, a firm's production output is seen as a function of factor inputs and an additive Hicks-neutral productivity term. Instead of specifying productivity as the outcome of a growth process common to all firms, we allow it to evolve differently between clusters. Upon its establishment, a firm makes a one-off discrete choice regarding cluster affiliation. This choice depends on the value the firm expects to derive from the cluster, which consists of its initial and expected future ability to produce output from factor inputs within the cluster and its—unobserved to the researcher—affinity for the cluster. This affinity can be driven by various factors, such as industry affiliation, innovative ability, trade potential, and beliefs regarding management practices, which are often not observable or readily available to researchers.

Building on a distributional assumption regarding a firm's affinity for productivity clusters, we model the probability of cluster affiliation per firm and cluster. This modeling does not require additional information on these clusters' drivers besides —standard in this literature— information on firm-level output and input factors. We then demonstrate that production function parameters are identified based on the degree of independence between timing decisions for factor input choices and *cluster-probability weighted* shocks to productivity (for an overview of currently prevalent identification schemes, see Ackerberg et al., 2007, 2015; Gandhi et al., 2020; De Loecker and Syverson, 2021). To take into account the simultaneity of factor input decisions along with cluster affiliation in this semi-parametrically defined environment, we deviate from the prevalent nonparametric generalized method of moments (GMM) and rely on limited information maximum likelihood (LIML) techniques by imposing a parametric assumption on the distribution of productivity. This allows us to obtain the production function parameters, the cluster-specific parameters of the productivity growth process, and the cluster affiliation probabilities within the same estimation procedure. Thus, by factoring latent heterogeneity into productivity, we can obtain "unbiased" estimates of productivity while reducing data requirements.[1]

We demonstrate the validity of the proposed methodology with a Monte Carlo analysis and, in turn, use detailed Belgian firm-level data to showcase its empirical applicability. First, we extend the Monte Carlo experiment by Ackerberg et al. (2015) to account for latent heterogeneity in productivity. We highlight the benefits of the proposed method relative to current estimators when firm-level drivers of heterogeneity are unobservable in the data or affected by measurement error. Unlike production function parameters and resulting productivity growth estimates, productivity differences between groups of firms are biased when latent heterogeneity is not controlled for. Second, we show that the proposed estimator provides economically sensible estimates when brought to the data.[2] We apply our estimator to firms in five Belgian manufacturing industries. We rely on data on firm-level revenue and input use from balance sheets and value-added tax (VAT) returns over the period 2008-2018, combined with a rich set of firm-level characteristics considered in the literature to be relevant for productivity growth, such as age, industry affiliation and participation in export, import, and foreign direct investment (FDI) activities. We find strong evidence of heterogeneity in productivity trends. Across all industries, we observe multiple firm clusters that differ in terms of productivity growth, as reflected by differences in the level of productivity, the magnitude of unexpected shocks to productivity, and the persistence of these shocks over time.

The results suggest cluster affiliation is positively associated with a firm's initial conditions, such as its initial productivity and factor input use. Any additional firm-level characteristics considered, such as age, export, import, or FDI status, can be associated with clusters but do not have strong explanatory power beyond the firm's initial conditions. This finding underlines the strength of the proposed estimation approach, whereby an unbiased identification of productivity does not necessarily require information on firm-level characteristics beyond output and input use. This aspect of the proposed estimator is key in settings with limited data availability and in the study of complex economic environments where it is inherently difficult to single out key drivers of productivity growth.

To empirically demonstrate how the proposed estimator differs from other commonly used

[1] By unbiased estimates, we mean estimates that focus on resolving the identification of productivity in the presence of latent heterogeneity, without taking a stance on other forms of biases that may arise when estimating productivity. See Van Beveren (2012); De Loecker and Goldberg (2014) and De Loecker and Syverson (2021) for an overview. The proposed estimation approach can easily be extended to incorporate existing solutions proposed in the literature to account for additional sources of bias liable to arise when estimating production functions.

[2] Notwithstanding its semi-parametric nature and the additional set of cluster-specific parameters, the empirical implementation of our estimation procedure remains computationally fast. This is attributable to the linear-in-parameters estimation problem for all cluster-specific sets of parameters. All relevant estimation code has been compiled in an easy-to-use R package which will be made publicly available upon publication.

methodologies that rely on information about firm-level characteristics, we focus on evaluating productivity differences between groups of firms.[3] To that end, we revisit a topic that has attracted the attention of multiple researchers and policymakers over the past years. Specifically, we examine the relative productivity advantage of exporting over non-exporting firms (the export premium), the evolution of this premium over time, and its contribution to aggregate productivity growth (see, for instance, Bernard and Bradford Jensen, 1999; Baldwin and Gu, 2003; Bernard et al., 2007; De Loecker, 2013; Garcia-Marin and Voigtländer, 2019; Gandhi et al., 2020). We demonstrate that export premia obtained from alternative productivity estimation methodologies can vary depending on the availability to the researcher of additional information on firm-level characteristics, such as age, export, import or FDI status. Thus, determining the contribution of exporting firms to aggregate productivity growth using these methods depends on the availability of firm-level information. In contrast, the proposed estimation approach delivers robust estimates of export premia and the contribution of exporting firms to aggregate productivity growth, regardless of the availability of additional information on firm-level characteristics in the dataset.

This paper is structured as follows. Section 2 situates the paper in the current literature. Next, the methodology is proposed and tested by means of a Monte Carlo analysis in Section 3. We subsequently apply the methodology to firm-level data in Section 4, before discussing the robustness of the results in Section 5. We end with a summary of the main contributions and opportunities for future research in Section 6.

2 Background and related literature

This paper builds on and extends the literature on structural production function estimation. Under certain assumptions regarding (i) the functional relationship between output and inputs; (ii) the timing of input decisions; and (iii) the evolution of productivity, the literature aims to identify the unobservable (to researchers) productivity based on various production function estimators (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015; Gandhi et al., 2020). We focus on the prevalent assumption of the productivity evolution 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 currently an unspecified functional form and η_{bt} represents an identically and independently distributed (i.i.d.) error term, also known as the innovation/productivity shock.

The predominant specification displayed in equation (1) assumes the evolution of productivity as a function of its lagged values and a random error term, which in turn assumes a homogeneous random productivity growth process for all firms. If additional firm-level characteristics, represented by the vector $\boldsymbol{e}_{bt}$, impact the evolution of productivity between clusters of firms,

[3] Productivity estimates are often used to evaluate productivity differences between groups of firms through a regression framework (Garcia-Marin and Voigtländer, 2019; Gandhi et al., 2020; Caliendo et al., 2020), or to evaluate the differential contribution of these groups to aggregate productivity growth through a decomposition framework (Collard-Wexler and De Loecker, 2015; Brandt et al., 2017).

they need to be specified:

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

Such a specification has been used to provide a wide array of empirical evidence on various firm-level drivers of productivity, such as innovation (Aw et al., 2011; Doraszelski and Jaumandreu, 2013; Bilir and Morales, 2020), trade (Amiti and Konings, 2007; Das et al., 2007; Kasahara and Rodrigue, 2008; De Loecker, 2013; Kasahara and Lapham, 2013; Merlevede and Theodorakopoulos, 2021), engagement in FDI (Javorcik, 2004; Blalock and Gertler, 2008), management practices (Bloom and Van Reenen, 2011; Caliendo et al., 2020; Rubens, 2020), technology (Harrigan et al., 2018), intangible transfers (Merlevede and Theodorakopoulos, 2020) and human capital (Van Beveren and Vanormelingen, 2014; Konings and Vanormelingen, 2015), among others. Table 1 contains a non-exhaustive list of studies identifying productivity drivers in the Markov process similar to equation (2).[4] The table shows the heterogeneity inherent in the use of firm-level characteristics as explanatory variables in the productivity process.

Allowing for heterogeneity as specified in equation (2) implies that the productivity process will be misspecified if a homogeneous productivity trend is imposed similar to that in equation (1). De Loecker (2013) shows that it is not possible to correctly identify the drivers of firm-level productivity without controlling for all relevant firm-level characteristics. This condition introduces heavy data requirements into the estimation procedure. Moreover, even if such heavy data requirements are met, some firm-level characteristics are expected to remain intrinsic and difficult to measure, e.g. managerial capacity and intangible capital (Haskel and Westlake, 2017). Finally, considering a large set of relevant and possibly highly collinear explanatory variables raises the need for sufficient variation in the data to circumvent potential collinearity issues and obtain precise point estimates. This seems infeasible in practice, especially for small segments or sectors of the economy with a limited number of observations available by design.

Table 1: Selected list of studies identifying productivity drivers in the Markov process

Study	Export	Import	R&D	FDI	Additional drivers
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

This paper proposes an extension of current production function estimation methods using FMM techniques which allow to capture unobservable heterogeneity originating from latent time-invariant firm-level determinants in the productivity process. Specifically, FMMs facilitate

[4]For the sake of completeness, this summary also considers studies in which firm-level characteristics are assumed as an input factor of production or ex-post associated with estimated productivity under an exogenous Markov process.

the specification of time-invariant clusters of firms with homogeneous growth processes defined as:

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

where the probability that firm b belongs to a certain cluster $s \in S$, $Pr(z_b^s|\cdot)$, and the total number of clusters (S) are determined by the data. In this way, differences between firm clusters can remain unspecified even if they are likely to be determined by various firm-level characteristics unobserved by the researchers.

We are not the first to propose a generalization of the Markov process specification to account for unobserved heterogeneity. Originally, Olley and Pakes (1996) envisioned a non-parametric specification of the productivity growth process but found it to be infeasible in practice (Olley and Pakes, 1996, footnote 23, p.1279). Dewitte et al. (2020) approximate productivity as a firm-specific fixed effect and a time trend, which can interact with each other in a non-parametric fashion, merely requiring a certain smoothness of productivity over time. However, if the smoothness requirement differs between unobserved groups of firms, this method could yield biased estimates (Li et al., 2016). Furthermore, Lee et al. (2019); Gandhi et al. (2020) and Ackerberg (2021) discuss the feasibility of allowing for firm-fixed effects in recent production function estimation techniques. As Gandhi et al. (2020) note, firm-fixed effects often lead to estimates of the capital coefficient that are unrealistically low and result in large standard errors. This is in line with Blundell and Bond (1998) who show that first-differenced production functions in a dynamic panel setup with a short time dimension perform poorly due to weak instruments.[5] As such, the methodology proposed in this paper generalizes the productivity trend semi-parametrically using FMMs. This flexible approach allows for cluster-specific constants that account for unobserved heterogeneity and maintains sufficient information to identify the production function parameters.

The advantages of FMMs have already been explored to consider technology-specific production function specifications. Van Biesebroeck (2003) uses the FMM framework to define a non-Hicks neutral value-added production function that differs between two technological clusters. Transition between clusters is possible but limited to technological upgrading. 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. (2015, 2017); Battisti et al. (2020) specify a non-Hicks neutral production technology with an a priori undefined number of clusters. Cluster affiliation is fixed over time. To control for the simultaneity problem, the authors rely on a parametrization of the first-order conditions with respect to factor inputs when estimating a value-added (Battisti et al., 2020) or gross-output (Kasahara et al., 2015, 2017) production function.

This paper shifts the focus from technology-specific production functions to generalizing current productivity-estimation techniques, allowing for latent heterogeneity in the evolution of

[5]To reduce such biases, one could further augment the estimation procedures borrowing from the 'system GMM' estimator developed by Blundell and Bond (1998) and outlined by Arellano and Bover (1995). For an application using the GNR methodology, see Merlevede and Theodorakopoulos (2021).

productivity. We generalize the estimation strategies for both value-added and gross-output production functions. To control for simultaneity problems, we rely on a LIML specification which does not require any additional assumptions regarding the first-order conditions of factor inputs beyond what is standard in the literature (Ackerberg et al., 2015; Gandhi et al., 2020). Moreover, we model the probability of belonging to a specific cluster in accordance with the behavioral framework specified below. This results in a mixture-of-experts specification (Gormley and Frühwirth-Schnatter, 2019) that improves cluster identification and allows ex-post inference to be drawn for each cluster, for instance, by evaluating the correlation between identified clusters and firm-level characteristics.

Finally, the advantages of FMMs in productivity have also been explored in the stochastic frontier literature (see, e.g., Beard et al., 1997; Orea and Kumbhakar, 2004; El-Gamal and Inanoglu, 2005; Greene, 2005). We build on the idea of controlling for the simultaneity problem when estimating a (non-FMM) production function using LIML from the stochastic frontier literature (for an overview, see Amsler et al., 2016). Compared to the literature, the structural production function estimation techniques impose less stringent functional form restrictions (Sickles and Zelenyuk, 2019).

3 Methodology

In this section, we first describe the behavioral framework considered. We then present our proposed production function identification and estimation strategy which we evaluate with a Monte Carlo exercise.

3.1 Behavioral framework

We assume a dynamic heterogeneous firms model with cluster-dependent uncertainty in Hicks-neutral productivity. The data consists of a (short) panel of firms $b = 1, \ldots, B$ over period $t = 0, \ldots, T$. Firms produce output Y_{bt} given a certain amount of capital K_{bt}, labor L_{bt} and materials M_{bt} in perfectly competitive output and input markets.[6] In each period t, firms have access to the a of information $\mathcal{I}_{bt}$ when making their production decisions. A generic input $X_{bt} \in \{K_{bt}, L_{bt}, M_{bt}\}$, is considered *non-flexible* if it is either predetermined $X_{bt} \in \mathcal{I}_{bt}$ or dynamic $X_{bt} = f^X(X_{bt-1})$.[7] Similarly, a *flexible* input is one that is neither predetermined $X_{bt} \notin \mathcal{I}_{bt}$ nor dynamic $X_{bt} \neq f^X(X_{bt-1})$ (Ackerberg et al., 2015; Gandhi et al., 2020). In this paper, we closely follow the popular structural production function estimation literature with at least one flexible input, assuming that capital and labor are non-flexible while materials is a flexible input.

[6] For the sake of simplicity, we limit the behavioral framework to the case of perfect competition. However, the proposed identification procedure solely affects the assumption about the Markov process of productivity and can, therefore, naturally be extended to settings that allow for imperfectly competitive output and input markets, as in Klette and Griliches (1996); De Loecker (2011); De Loecker et al. (2016); Rubens (2021), and Blum et al. (2021).

[7] Throughout this paper, we differentiate between functional forms $f(\cdot)$ by indexing them with input factors of interest x, i.e. $f^x(\cdot)$.

The relationship between outputs and inputs is expressed as follows:

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

where lowercase variables indicate logarithmic values of uppercase variables. $f^{klm}(\cdot)$ represents the production function explaining the variability in firm-level output, along with two additive terms. On the one hand, $\varepsilon_{bt} \notin \mathcal{I}_{bt}$ represents an ex-post shock to production and possible classical measurement error that does not affect future output. On the other hand, $\omega_{bt} \in \mathcal{I}_{bt}$ represents Hicks-neutral total factor productivity (TFP) that is known to the firm before making its period t decisions. Concretely, ω_{bt} "might represent variables such as the managerial ability of a firm, expected downtime due to machine breakdown, expected defect rates in a manufacturing process, soil quality, or the expected rainfall at a particular farm's location", while ε_{bt} "might represent deviations from expected breakdown, defect, or rainfall amounts in a given year" (Ackerberg et al., 2015, p.2414).

In contrast to the existing literature, we generalize the productivity component ω_{bt} to evolve following a *cluster-dependent* first-order Markov process

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

where each firm b belongs to a certain cluster $s = 1, \ldots, S$, indicated by $z_b^i = \mathbb{I}_b\left(s = i\right), \forall i = 1, \ldots, S$ which affects its productivity over time. $\mathbb{I}(\cdot)$ is an indicator function and $p(\cdot)$ represents the probability density function. We assume that the total number of clusters S is exogenously determined.[8] Note that this generalized Markov process nests the single-cluster specification (1) commonly employed in the literature, i.e., when $S = 1$.

At entry, $t = 0$,[9] the firm makes a net present value comparison between clusters and chooses the cluster to belong to from the next period onward that will result in the highest discounted profits, taking expectations and the costs of cluster affiliation into account.[10] This results in an optimal decision rule for cluster affiliation:

$$\begin{aligned} z_b^*\left(K_{b0}, L_{b0}, e^{\omega_{b0}}, \xi\right) = \arg\max_{z_b^s} \Bigg(& \pi_0\left(K_{b0}, L_{b0}, e^{\omega_{b0}}\right) + \xi(z_b^s) + \\ & E_\omega\left[\sum_{t=1}^{T} \beta^{t-1}\pi_t\left(K_{bt}, L_{bt}, e^{\omega_{bt}}, z_b^s\right)\right]\Bigg) \end{aligned} \tag{6}$$

[8]We do not attempt to impose a behavioral framework on the number of clusters as there is currently a lack of understanding or too little information on the exact drivers of clusters in the literature.

[9]In the empirical application below, we broadly interpret entry as the first year a firm is observed in the dataset. The idea is to economically justify the specification of dependence between a firm's initial state and the latent cluster affiliation indicator z_b.

[10]The idea of fixed cluster membership over time is not at odds with the Belgian firm-level data we use for our empirical application in the next section. Over ten years, 100% of the Belgian firms do not change their location, 91.9% do not change their industry affiliation, 81.8% do not change their export status, 72.6% do not change their import status, and 98.1% do not change their FDI status. Moreover, imminent changes to firm status can be expected to be related to their initial conditions and result in appropriately differentiated clusters.

where $\beta \in (0,1)$ is the discount factor, $\pi_t(\cdot)$ is the profit function giving current-period profits as a function of the vector of state variables, and $\xi(\cdot)$ is a choice-specific i.i.d. variable that captures the affinity of a firm for a particular cluster.[11] This affinity is observed by the firm but not by the econometrician. Therefore, we specify the probability density function for ξ by $f^{\xi}(\xi)$ and obtain the probability of a firm choosing a cluster s conditional on initial values of capital, labor and productivity by integrating the decision rule over the regions of ξ for which $z_b^*(\cdot) = z_b^s$ (Arcidiacono et al., 2011):

$$Pr(z_b^s|K_{b0}, L_{b0}, e^{\omega_{b0}}) = \int \mathbb{I}\left[z_b^*\left(K_{b0}, L_{b0}, e^{\omega_{b0}}\right) = z_b^s\right] f^{\xi}(\xi) d\xi. \tag{7}$$

This equation is key to identifying cluster affiliation in our empirical strategy. As we demonstrate below, it allows us to resolve the initial conditions problem, i.e., the problem of how to factor in dependence between a firm's initial state and latent cluster affiliation z_b. It should be noted how closely the specification relates to Olley and Pakes (1996), who specify a firm's unobserved exit policy rule as a function of observed state variables. While they rely on their specification to model the probability of firm exit and control for selection bias, we exploit the information available in the observed initial state variables to identify unobserved cluster affiliation.[12]

3.2 Production function estimation

Identification of the parameters that define the production function specified in equation (4) is burdened by a simultaneity problem. Specifically, firm-level input choices depend on and thus correlate with the unobserved productivity term, i.e., $E\left[\omega_{bt}|k_{bt}, l_{bt}, m_{bt}\right] \neq 0$. This dependence renders ordinary least squares (OLS) or nonlinear least squares (NLS) estimates of output elasticities inconsistent. Therefore, alternative identification strategies have been developed, usually consisting of two stages.

In the *first stage*, the ex-post production term (ε_{bt}) and the contribution of the flexible input factors are separated from output in the main estimating equation (4). Different methods exist to do so. For instance, Ackerberg et al. (2015) rely on a value-added production function and assume proportionality of the flexible production factor m_{bt} to value added. They then use the flexible production factor as a control variable for productivity to identify the ex-post shock to production and classical measurement error term ε_{bt}. On the other hand, Gandhi et al. (2020) build on the first-order conditions of the flexible production factor m_{bt} to jointly identify the ex-post production term and the output elasticity of the flexible input from a gross output production function.[13]

Regardless of the production function estimation methodology used, the first stage results in an equation of this form:

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

[11] For instance, the affinity could translate into a lower fixed cost of cluster affiliation.
[12] Note that we can also factor into our model an exit rule as in OP to control for the selection problem.
[13] See Appendix B for a detailed description of the first stage of both methodologies.

where ϕ_{bt} represents the remaining output variation after netting out the estimates of the first stage ex-post shocks to production and, for the case of a gross-output production function, the output contribution of the flexible production factor. Up to this point, the steps taken are standard in the literature.

The *second stage* allows the identification of output elasticities of non-flexible inputs. It relies on the assumed Markov process of productivity in (5) to replace the unobserved productivity term ω_{bt} from (8) as a function of observables and production function parameters. The novel part of our methodology is that we generalize the productivity evolution process to explicitly depend on the fixed cluster affiliation of a firm through the cluster affiliation indicator z_b^s:[14]

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

Identification of the production function coefficients relies on the independence of cluster-specific 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):

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

If $S = 1$, the moment conditions simplify to those commonly used in the production function literature, and equation (9) can be estimated with GMM (Ackerberg et al., 2015; Gandhi et al., 2020). When $S > 1$, the moment conditions contain a cluster affiliation indicator that is latent to the researcher. Therefore, to estimate equation (9), we employ a finite mixture specification for the evolution of productivity and rely on the expectation-maximization algorithm to jointly estimate the production function parameters, the cluster-specific parameters of the productivity process and the unobserved cluster affiliation probabilities. To specify cluster probabilities without additional information, we rely on the behavioral framework presented in Section 3.1 and a parametric functional form assumption about the firm's affinity for specific clusters. For endogeneity concerns regarding labor input decisions (Doraszelski and Jaumandreu, 2013; Ackerberg et al., 2015), we condition on lagged labor values to instrument for current labor using a reduced form instrumental equation. This results in a LIML specification.

Our framework requires three parametric assumptions compared to the non-parametric GMM estimator. First, we assume cluster-specific log-productivity growth is log-normal. Second, we model endogenous labor as a reduced-form function of exogenous instruments and a normally distributed error term. Third, we assume the unobserved choice-specific i.i.d. variable $\xi(z_b^s)$ follows a type-1 extreme value distribution.

[14]This generalization of the Markov process is consistent with the first-stage estimation procedures discussed. In particular, the first stage relies on a flexible production factor unaffected by differences in the expectations of future productivity shocks between groups of firms (Ackerberg, 2021). See Section 6 for a discussion of the adequacy of this assumption.

Below, we first specify the observed likelihood by conditioning on lagged labor values. We then determine the complete likelihood that accounts for cluster affiliation before we present the estimation algorithm and discuss model selection algorithms over the total number of clusters S.

3.2.1 Observed likelihood

We parameterize equation (9) assuming that productivity follows a Gaussian mixture. This specification is in line with Dewitte et al. (2022) who find that firm-size distribution is best represented by a finite mixture of log-normals.[15] As a result, the probability of observing a single observation becomes:

$$p^o(\phi_{bt}|k_{bt}, l_{bt}, \phi_{bt-1}, l_{bt-1}, k_{bt-1}, z_b^s; \boldsymbol{\beta}, \boldsymbol{\alpha}^s, \sigma_\eta^s) = \frac{1}{\sigma_\eta^s}\varphi\left(\frac{f^{kl}\left(k_{bt}, l_{bt}; \boldsymbol{\beta}\right) - g(\phi_{bt-1}, k_{bt-1}, l_{bt-1}; \boldsymbol{\beta}, \boldsymbol{\alpha^s})}{\sigma_\eta^s}\right), \tag{11}$$

where $\varphi(\cdot)$ represents the standard-normal density.

It is possible that labor 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 needs to be taken into account when specifying the observed likelihood. Therefore, we specify the following reduced-form equation for endogenous labor with exogenous instruments $k_{bt}, l_{bt-1}, \phi_{bt-1}$ and a normally distributed error term ($\zeta_{bt} \sim \mathcal{N}\left(0, \sigma_\zeta\right)$), such that:[16]

$$l_{bt} = \delta_0^s + \delta_1 k_{bt} + \delta_2^s \phi_{bt-1} + \delta_3^s k_{bt-1} + \delta_4^s l_{bt-1} + \zeta_{bt}^s. \tag{12}$$

It should be noted that we specify the current capital coefficient 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 attains a bivariate normal specification:[17]

$$p^o(\phi_{bt}, l_{bt}|k_{bt}, \phi_{bt-1}, k_{bt-1}, l_{bt-1}, z_b^s; \underbrace{\boldsymbol{\beta}, \boldsymbol{\alpha}^s, \boldsymbol{\delta}^s, \boldsymbol{\Sigma}^s}_{\equiv \boldsymbol{\theta}^s}) = \frac{e^{-\frac{1}{2}(\boldsymbol{\epsilon}^s)^T(\boldsymbol{\Sigma}^s)^{-1}(\boldsymbol{\epsilon}^s)}}{\sqrt{(2\pi)^2|\boldsymbol{\Sigma}^s|}}, \tag{13}$$

[15] Aside from empirical evidence, two arguments favor the (log-)normal specification of productivity. First, from the perspective of overall fit, a mixture of normal distributions with sufficient components is shown to be able to approach all distributions (McLachlan and Peel, 2000). This argument implies, however, that the number of mixtures does not necessarily coincide with the number of clusters in the data. Second, from a generative perspective for individual components, the normal distribution is the realization of applying the central limit theorem, whereby firm productivity is approximately normally distributed if it is the sum of many independent random variables. This corresponds to the multi-dimensional definition of productivity when accounting for the product dimension (Bernard et al., 2009) or uncertainty in demand and/or supply (see, for instance, De Loecker, 2011; Bas et al., 2017; Gandhi et al., 2020).

[16] We closely follow Ackerberg et al. (2015), and Gandhi et al. (2020), and rely on an exactly identified case with a one-period lagged instrument for labor as our main specification. Including additional instruments is feasible with this methodology.

[17] Similarly, Doraszelski and Jaumandreu (2013) rely on a system of simultaneous equations to estimate productivity under endogeneity.

where $\boldsymbol{\epsilon}^s = \begin{bmatrix} \phi_{bt} - f^{kl}\left(k_{bt}, l_{bt}; \boldsymbol{\beta}\right) - g(\phi_{bt-1}, k_{bt-1}, l_{bt-1}; \boldsymbol{\beta}, \boldsymbol{\alpha^s}) \\ l_{bt} - \delta_0^s - \delta_1 k_{bt} - \delta_2^s \phi_{bt-1} - \delta_3^s k_{bt-1} - \delta_4^s l_{bt-1} \end{bmatrix}$ and $\boldsymbol{\Sigma}^s = \begin{bmatrix} (\sigma_\eta^s)^2 & \sigma_{\eta,\zeta}^s \\ \sigma_{\eta,\zeta}^s & (\sigma_\zeta^s)^2 \end{bmatrix}$.

To determine the probability of observing the complete data series, it is necessary to model the conditional probability of belonging to a specific cluster from period one onwards (see equation (7)). We rely on the behavioral framework presented in Section 3.1 and a parametric functional form assumption on the firm's affinity for specific clusters to specify the probabilities of unobserved cluster affiliation. We specify these probabilities as a function of the observed initial conditions based only on firm-level information already available for the production function estimation. Assuming the unobserved choice-specific i.i.d. variable $\xi(z_b^s)$ follows a type-1 extreme value distribution and relying on the reduced form of the conditional choice probability (7), this probability can be modeled as follows (McFadden, 1973; Arcidiacono et al., 2011):[18,19]

$$Pr(z_b^s | k_{b0}, l_{b0}, \omega_{b0}; \boldsymbol{\gamma}^1, \ldots, \boldsymbol{\gamma}^s) = \frac{e^{\gamma_0^s + \gamma_k^s k_{b0} + \gamma_l^s l_{b0} + \gamma_\omega^s \omega_{b0}}}{\sum_{i=1}^S e^{\gamma_0^i + \gamma_k^i k_{b0} + \gamma_l^i l_{b0} + \gamma_\omega^i \omega_{b0}}}, \qquad \forall s = 1, \ldots, S. \tag{14}$$

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

$$\begin{aligned} p^o(\boldsymbol{\phi}, \boldsymbol{l}; \underbrace{\{\boldsymbol{\gamma}^1, \ldots, \boldsymbol{\gamma}^S, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^S\}}_{\equiv \boldsymbol{\Theta}}) = \\ \prod_{b=1}^B \sum_{s=1}^S Pr(z_b^s | k_{b0}, l_{b0}, \omega_{b0}; \boldsymbol{\gamma}^s) \prod_{t=1}^T p(\phi_{bt}, l_{bt} | k_{bt}, \phi_{bt-1}, k_{bt-1}, l_{bt-1}, z_b^s; \boldsymbol{\theta}^s). \end{aligned} \tag{15}$$

This specification naturally accounts for the initial conditions problem, the dependence between the initial dependent variables (ϕ_{b0}, l_{b0}), and the latent cluster affiliation indicator z_b^s. Following (Wooldridge, 2005; Frühwirth-Schnatter et al., 2012), the specification above relies on the factorization of the joint distribution for ϕ_{b0}, l_{b0} and z_b^s into a model for z_b^s conditional on the initial state variables and a marginal model for the initial dependent variables: $p(\phi_{b0}, l_{b0}, z_b^s) = Pr(z_b^s | k_{b0}, l_{b0}, \omega_{b0}) p(\phi_{b0}, l_{b0})$, where we rely on the functional relation between ϕ_{b0} and the initial state variables $(k_{b0}, l_{b0}, \omega_{b0})$. Moreover, as the marginal model for the initial dependent variables is cluster-independent, it does not need to be explicitly specified because it cancels out from all posterior distributions specified during the estimation procedure discussed below.

3.2.2 Complete log-likelihood

From (15), we can specify the observed (o) log-likelihood as:

$$\mathcal{L}^o(\boldsymbol{\Theta}) = \sum_{b=1}^B \sum_{s=1}^S log\left(Pr(z_b^s | k_{b0}, l_{b0}, \omega_{b0}; \boldsymbol{\gamma}^s) \prod_{t=1}^T p(\phi_{bt}, l_{bt} | k_{bt}, \phi_{bt-1}, k_{bt-1}, l_{bt-1}, z_b^s; \boldsymbol{\theta}^s) \right), \tag{16}$$

[18]For the sake of simplicity, we rely on a reduced form optimal decision rule. See Arcidiacono and Miller (2011) for a discussion of the estimation of dynamic discrete choice models with unobserved heterogeneity.

[19]The logit specification relies on the independence of irrelevant alternatives assumption. It follows from the specification in (6) that this assumption is satisfied.

which does not account for the unobserved cluster affiliation. Additionally, when accounting for cluster affiliation (z_b^s), the complete (c) log-likelihood becomes:

$$\mathcal{L}^c\left(\boldsymbol{\Theta},\boldsymbol{z}\right)=\sum_{b=1}^{B}\sum_{s=1}^{S}z_b^s log\left(Pr(z_b^s|k_{b0},l_{b0},\omega_{b0};\boldsymbol{\gamma}^s)\prod_{t=1}^{T}p(\phi_{bt},l_{bt}|k_{bt},\phi_{bt-1},k_{bt-1},l_{bt-1},z_b^s;\boldsymbol{\theta}^s)\right), \tag{17}$$

which forms the basis for our estimation procedure.

3.2.3 Estimation procedure

To estimate the parameters of interest in (17), we rely 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. The parameter values in the j^{th} iteration are represented as $(\boldsymbol{\Theta})^j\equiv\left\{(\boldsymbol{\gamma}^1)^j,\ldots,(\boldsymbol{\gamma}^S)^j,(\boldsymbol{\theta}^1)^j,\ldots,(\boldsymbol{\theta}^S)^j\right\}$. Each iteration includes the following steps:

1. Use the current-iteration starting values for the parameters $(\boldsymbol{\Theta})^j$ and approximate cluster affiliation with the posterior conditional probability obtained from Bayes' theorem:

$$\hat{z}_b^s=Pr\left(z_b^s|\boldsymbol{k}_b,\boldsymbol{l}_b,\boldsymbol{\phi}_b;(\boldsymbol{\Theta})^j\right)=\frac{Pr\left(z_b^s|k_{b0},l_{b0},\omega_{b0};(\boldsymbol{\gamma}^s)^j\right)\prod_{t=1}^{T}p(\phi_{bt},l_{bt}|k_{bt},\phi_{bt-1},k_{bt-1},l_{bt-1},z_b^s;\boldsymbol{\theta}^s\Big)}{p^o\left(\boldsymbol{\phi},\boldsymbol{l};(\boldsymbol{\Theta})^j\right)}. \tag{18}$$

2. Use these cluster affiliation estimates to estimate the parameters $(\boldsymbol{\Theta})^{j+1}$ by maximizing the complete log-likelihood $\mathcal{L}^c\left((\boldsymbol{\Theta})^{j+1},\hat{\boldsymbol{z}}\right)$:

$$\text{(i)}\max_{(\boldsymbol{\theta})^{j+1}}\sum_{b=1}^{B}\sum_{s=1}^{S}Pr(z_b^s|\boldsymbol{k}_b,\boldsymbol{l}_b,\boldsymbol{\phi}_b;(\boldsymbol{\Theta})^j)log\left(\prod_{t=1}^{T}p(\phi_{bt},l_{bt}|k_{bt},\phi_{bt-1},k_{bt-1},l_{bt-1},z_b^s;(\boldsymbol{\theta}^s))^{j+1}\right);$$

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

This iterative process continues until there is relative stability between iterations j and $j+1$ in terms of the observed log-likelihood (16) (see Appendix B for a detailed description of the estimation methodology).

3.3 Comparison with alternative identification strategies

It is informative to discuss the prevalent alternative identification strategies for the production function vis-á-vis the existence of clusters in the productivity process. Based on the existing literature, we consider both (i) a unitary cluster affiliation and (ii) a deterministic cluster affiliation and discuss how (iii) the random cluster affiliation proposed in this paper nests the cluster affiliation specification of (i) and mimics specification (ii).

(i) Unitary cluster affiliation. Assuming $S=1$, (9) takes the well-known form:

$$\phi_{bt}=f^{kl}\left(k_{bt},l_{bt}\right)+g^1\left(\phi_{bt-1}-f^{kl}\left(k_{bt-1},l_{bt-1}\right)\right)+\eta_{bt}^1. \tag{19}$$

While this is the specification commonly used in the literature, it is misspecified if there is at least one group of firms that evolves differently over time, i.e. $S > 1$:[20]

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

(ii) Deterministic cluster affiliation. 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 $N = S$, then:

$$\hat{z}_b^s = \mathbb{I}\left(\boldsymbol{e}_b = e_b^s\right), \qquad \forall s = 1, \ldots, S, \tag{20}$$

and therefore:

$$E\left[\sum_{s=1}^{S} \mathbb{I}\left(\boldsymbol{e}_b = e_b^s\right) \eta_{bt}^s \middle| k_{bt}, l_{bt(-1)}, \phi_{bt-1}\right] = 0. \tag{21}$$

Usually, to the researcher, it is a priori unknown whether these variables determine cluster affiliation and whether $N \neq S$. If $N < S$, the process is misspecified, the likelihood of which is present as discussed in Section 2 and emphasized in Table 1. Moreover, measurement error in the categorical variables leads to cluster misallocation and, thus, a biased estimator.

(iii) Random cluster affiliation. The identification and estimation strategy proposed in this paper models the unobserved cluster affiliation z_b^s as a random variable with its probability determined solely from readily available information:

$$E\left[\sum_{s=1}^{S} Pr(z_b^s | \boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \hat{\boldsymbol{\Theta}}) \eta_{bt}^s \middle| k_{bt}, l_{bt(-1)}, \phi_{bt-1}\right] = 0. \tag{22}$$

Overall, this approach has the advantage of identifying the production function parameters without prior knowledge of firm cluster affiliation. In the unlikely event prior information regarding cluster affiliation ($\boldsymbol{e}_b$) is available, the proposed approach is as good as the deterministic approach (ii).[21] Finally, reliance on a random specification implies that this approach

[20] Under a Cobb-Douglas production function and an AR(1) productivity process with two firm-clusters ($S = 2$):

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

However, if we assume a unitary cluster affiliation ($S = 1$), the specification becomes:

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

and thus, if $\alpha_{0,1}^{1,2} \neq 0$, then the omitted cluster-indicator is by construction correlated with the remaining explanatory variables and will bias the estimated coefficients (for an in-depth discussion, see De Loecker, 2013).

[21] Assume there are two clusters with a priori known cluster affiliation, i.e., we observe the indicator variable $\mathbb{I}_b\left[s = 1\right]$, then:

$$ln\frac{Pr(z_{bt}^1)}{Pr(z_{bt}^2)} = \gamma_0^1 + \gamma_k^1 k_{b0} + \gamma_l^1 l_{b0} + \gamma_\omega^1 \omega_{b0} + \gamma_1^1 \mathbb{I}_b\left[s = 1\right],$$

with the prior probabilities approximately equal to unity:

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

also allows for measurement error in the categorical variables (see the Monte Carlo simulation below).

The advantages of the proposed approach are obtained through functional form restrictions. Compared with the non-parametric GMM estimator, our framework requires imposing three parametric assumptions. First, we assume cluster-specific log-productivity growth is log-normal. Second, we model endogenous labor as a reduced-form function of exogenous instruments and a normally distributed error term. Third, we assume the unobserved choice-specific i.i.d. variable $\xi(z_b^s)$ follows a type-1 extreme value distribution.

3.4 Model selection

While the number of clusters S is assumed to be an exogenous variable in our economic model (see Section 3.1), we allow the data to determine this number. Testing the order of the finite mixture using likelihood ratio tests is difficult and rarely done, as regularity conditions that ensure a standard asymptotic distribution for the maximum likelihood estimates do not hold (Celeux et al., 2018). Therefore, we approach this step as a model selection problem, in which we estimate the model for several clusters and rely on evaluation criteria to determine the "true" number of clusters (Celeux et al., 2018). We rely on two evaluation criteria: the Bayesian information criterion (BIC) and the integrated complete-data likelihood Bayesian information criterion (ICLbic). If these evaluation criteria prefer a multi-cluster over a single-cluster model specification, we interpret this as a rejection of the homogeneity assumption for the productivity growth process.

The BIC is based on penalizing the observed log-likelihood function (16) proportional to the number of free parameters (np) in the model, such that:

$$BIC(S) = -2\mathcal{L}^o(\hat{\boldsymbol{\Theta}}) + nplog(BT). \tag{23}$$

The optimal model minimizes the BIC criterion over S. As such, it favors parsimonious models and is consistent in selecting the number of mixture components when the mixture model is used to estimate a density (Celeux et al., 2018).

One limitation is that the BIC does not consider the purpose of the modeling. It does not account for the usefulness of additional clusters when assessing S, i.e., how well separated the different clusters are. Clusters are well separated if $\hat{z}_b^s$ is close to 1 for one component and close to 0 for all other components. Therefore, as an alternative criterion, we consider ICLbic, which selects S such that the resulting mixture model leads to a clustering of the data with the largest evidence base (Biernacki et al., 2000):

$$ICLbic(S) = -2\left(\mathcal{L}^o(\hat{\boldsymbol{\Theta}}) + \left[\sum_{s=1}^{S}\sum_{b=1}^{B}\sum_{t=1}^{T} Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \hat{\boldsymbol{\Theta}})log\left(Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \hat{\boldsymbol{\Theta}})\right)\right]\right) + \frac{np}{2}log(BT). \tag{24}$$

This prior information on cluster affiliation is validated by the data and results in a close to perfect identification of the posterior probability of cluster affiliation $\hat{z}_b^1 = Pr(z_b^1|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b, \mathbb{I}_b\left[s=1\right]; \hat{\boldsymbol{\Theta}}) \approx 1$.

The optimal model maximizes the ICLbic criterion over S. For example, if the mixture components are well separated for a given S, then the term in brackets above tends to define a clear partition of the dataset. If this is the case, the term is close to 0. On the other hand, if the mixture components are poorly separated, the term takes values larger than zero. Due to this additional term, the ICLbic criterion favors values of S that give rise to partitions of the data with the strongest evidence base. In practice, ICLbic appears to provide a stable and reliable selection of S for real data sets (Celeux et al., 2018).

3.5 Monte Carlo

We conduct a Monte Carlo (MC) exercise to evaluate the estimator's performance. The focus is on the estimator's ability to recover unobserved heterogeneity in the productivity distribution. The setup of our MC analysis closely mimics that of Ackerberg et al. (2015), which builds on Syverson (2001) and Van Biesebroeck (2007). The key deviation from Ackerberg et al. (2015) is in the specification of the productivity Markov process, which is assumed to differ between firm clusters.[22] Specifically, productivity is assumed to follow a finite mixture AR(1) process with two clusters ($S = 2$):

$$\omega_{bt} = \sum_{s=1}^{2} z_b^s \left[\alpha_0^s + \alpha_1^s \omega_{bt-1} + \eta_{bt}^s\right], \tag{25}$$

with 800 firms exogenously assigned to cluster one ($s = 1$) with probability $Pr(z_b^1) = 0.8$, and 200 firms to cluster two ($s = 2$) with probability $Pr(z_b^2) = 0.2$. Furthermore, we follow Ackerberg et al. (2015) in assuming a normal distribution for the cluster-specific productivity shocks $\eta_{bt}^s \sim \mathcal{N}\left(0, \sigma_\eta^s\right)$.

Firms make optimal capital investment choices to maximize the expected (discounted) value of future profits under convex capital adjustment costs such that the period t capital stock (K_{bt}) is determined by investment at $t-1$, i.e., $K_{bt} = (1-\delta)K_{bt-1} + I_{bt-1}$. Material inputs ($M_{bt}$) are chosen at t, while labor input (L_{bt}) is chosen either at t or at $t-i$ (in the latter case, labor is chosen with only knowledge of $e^{\omega_{b,t-i}}$ where $i \leq 1$, not $e^{\omega_{bt}}$) (Ackerberg et al., 2015). The production function is assumed Leontief in (and proportional to) materials, such that:

$$Y_{bt} = min\left\{K_{bt}^{\beta_k} L_{bt}^{\beta_l} e^{\omega_{bt}}, \beta_m M_{bt}\right\} e^{\varepsilon_{bt}}, \tag{26}$$

where the true values of the output elasticity for each input are $\beta_k = 0.4, \beta_l = 0.6$, and $\beta_m = 1$. This assumes a Leontief production technology, which results in the following value-added production function:

$$\frac{y_{bt}}{m_{bt}} = \beta_k k_{bt} + \beta_l l_{bt} + \omega_{bt} + \varepsilon_{bt}. \tag{27}$$

We next specify four different data-generating processes (DGPs). The first DGP (DGP1) assumes no difference in the parameters of productivity evolution across clusters, where $\alpha_0^1 = \alpha_0^2 = 1$, $\alpha_1^1 = \alpha_1^2 = 0.7$ and $\sigma_\omega^1 = \sigma_\omega^2 = 0.3$. This specification is equivalent to the case without latent heterogeneity and identical to the DGP in Ackerberg et al. (2015). As such, the MC analysis

[22] See Appendix C for a complete description of the MC simulation.

on the DGP1 allows us to evaluate the appropriateness of the LIML vis-à-vis the traditional GMM specification.

The second DGP (DGP2) introduces latent heterogeneity through differences in the productivity evolution between clusters of firms. This allows us to evaluate the ability of the proposed methodology to identify unobserved clusters between which the evolution of productivity differs. We specify $\alpha_0^1 = 1$ and $\alpha_0^2 = 0.8$, $\alpha_1^1 = 0.7$ and $\alpha_1^2 = 0.77$, and $\sigma_\omega^1 = 0.3$ while $\sigma_\omega^2 = 0.39$. Overall, this specification results in an approximate 14.5% stationary average productivity advantage for the second cluster.[23]

In the third DGP (DGP3), we assume that cluster affiliation is observed by the researcher such that the prior probability of cluster affiliation can be identified as:

$$Pr(z_b^s|k_{b0}, l_{b0}, \omega_{b0}; \boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2) = \frac{e^{\gamma_0^s+\gamma_k^s k_{b0}+\gamma_l^s l_{b0}+\gamma_\omega^s \omega_{b0}+\gamma_{cluster}^s \mathbb{I}(s=s)}}{\sum_{i=1}^S e^{\gamma_0^i+\gamma_k^i k_{b0}+\gamma_l^i l_{b0}+\gamma_\omega^i \omega_{b0}+\gamma_{cluster}^i \mathbb{I}(i=s)}}, \qquad \forall s = 1, \ldots, S. \quad (28)$$

Finally, the fourth DGP (DGP4) builds on DGP3 but assumes that 10% of firms are misclassified in clusters. This is a more realistic scenario for researchers and allows a comparison of the deterministic approach and our proposed random approach to cluster affiliation.

To estimate (27) for all DGPs, we follow the Ackerberg et al. (2015) estimation approach for the first stage. The second stage differentiates between identification strategies. Specifically, we follow the discussion in Section 3.3 and estimate: (i) a unitary cluster affiliation according to (Ackerberg et al., 2015) (Uni. GMM), (ii) a deterministic cluster affiliation where the cluster identification variable is observed (Det. GMM) and (iii) our proposed estimation approach with random one-cluster ($S = 1$) and two-cluster ($S = 2$) affiliation imposed, named as 1-comp. LIML and 2-comp. LIML, respectively.[24]

Table 2 displays the results of the MC analysis. Focusing on the evolution of productivity, we display the normalized mean squared Error (MSE)[25] of the Markov process parameters α_0, α_1, σ_η for both clusters, along with the NMSE of the average share-weighted productivity growth $\bar{\Omega} = \sum_{t=1}^T \sum_{b=1}^B \frac{share_{bt}\omega_{bt}}{T}$, where $share_{bt} = \frac{y_{bt}}{\sum_{b=1}^B y_{bt}}$, and of the average cluster productivity premium $\bar{\omega}^2 - \bar{\omega}^1$ where $\bar{\omega}^s = \sum_{b\in s} \frac{\omega_{bt}}{T\sum \mathbb{I}(b\in s)}$. The NMSE allows to evaluate the bias and variance of the estimator in one statistic. It should be noted that in the case of the 2-comp. LIML, we rely on the model-identified rather than the imposed cluster affiliation to calculate these statistics.

From the results for a single cluster (DGP1), we observe that the LIML identification procedure accurately estimates Markov parameters, similar to the prevalent Uni. GMM. Allowing for multiple clusters in a single-cluster environment results in efficiency losses, as shown in the Det. GMM and 2-comp. LIML estimates, with an over-estimated average cluster productivity premium for the 2-comp. LIML.

[23] $\frac{0.8}{1-0.77} - \frac{1}{1-0.7} \approx 0.145$

[24] For all estimators, starting values for the parameters are set to $\beta_k = 0.3$ and $\beta_l = 0.7$.

[25] $NMSE = \frac{\sum_i(\hat{x}_i - x_i)^2}{N\sum_i(\hat{x}_i)^2}$ for each true coefficient x and its estimate $\hat{x}$ over each Monte Carlo iteration i.

Table 2: Monte Carlo results

Methodology	α_0^1	α_1^1	σ_η^1	α_0^2	α_1^2	σ_η^2	$\bar{\Omega}$	$\bar{\omega}^2 - \bar{\omega}^1$
			DGP1 - No Heterogeneity					
Uni. GMM	0.00113	0.00019	0.00018	-	-	-	0.00063	0.00127
Det. GMM	0.00126	0.00020	0.00018	0.00382	0.00072	0.00018	0.00063	0.00128
1-comp. LIML	0.00107	0.00019	0.00018	-	-	-	0.00061	0.00119
2-comp. LIML	0.00441	0.00067	0.00065	0.03855	0.00656	0.00227	0.00095	1.04957
			DGP2 - Latent Heterogeneity					
Uni. GMM	0.04452	0.00051	0.00153	-	-	-	0.00998	2.44013
Det. GMM	0.00172	0.00022	0.00117	0.00618	0.00047	0.01583	0.00105	0.03256
1-comp. LIML	0.04466	0.00051	0.00152	-	-	-	0.01003	2.45750
2-comp. LIML	0.00225	0.00025	0.00024	0.00906	0.00070	0.00054	0.00155	0.10215
			DGP3 - Observed Heterogeneity					
Uni. GMM	0.04020	0.00034	0.00151	-	-	-	0.00654	2.37277
Det. GMM	0.00148	0.00018	0.00117	0.00629	0.00054	0.01576	0.00064	0.03064
1-comp. LIML	0.04013	0.00034	0.00149	-	-	-	0.00652	2.35942
2-comp. LIML	0.00171	0.00024	0.00023	0.00735	0.00067	0.00063	0.00068	0.04285
			DGP4 - Measurement error in Observed Heterogeneity					
Uni. GMM	0.04355	0.00042	0.00167	-	-	-	0.00963	2.08070
Det. GMM	0.00953	0.00022	0.00148	0.00440	0.00127	0.01442	0.00526	0.44151
1-comp. LIML	0.04346	0.00041	0.00165	-	-	-	0.00969	2.06406
2-comp. LIML	0.00198	0.00023	0.00020	0.00595	0.00053	0.00044	0.00116	0.19822

Notes: Results display the normalized mean squared error, accommodating the estimator's bias and variance, of the estimates obtained across 100 Monte Carlo iterations. α_0^s, α_1^s, and σ_η^s represent the cluster-specific constant, auto-regressive parameter, and standard deviation of the productivity shock, respectively. $\bar{\Omega} = \sum_{t=1}^{T} \sum_{b=1}^{B} \frac{share_{bt}\omega_{bt}}{T}$, where $share_{bt} = \frac{y_{bt}}{\sum_{b=1}^{B} y_{bt}}$, represents the average share-weighted productivity growth of the complete data. $\bar{\omega}^2 - \bar{\omega}^1$ represents the average cluster productivity premium with $\bar{\omega}^s = \sum_{b \in s} \frac{\omega_{bt}}{T \sum \mathbb{I}(b \in s)}$. The true coefficients in DGP1 are as follows: $\alpha_0^1 = \alpha_0^2 = 1$, $\alpha_1^1 = \alpha_1^2 = 0.7$ and $\sigma_\omega^1 = \sigma_\omega^2 = 0.3$. The true coefficients in DGP2-4 are as follows: $\alpha_0^1 = 1$, $\alpha_0^2 = 0.8$ $\alpha_1^1 = 0.7$, $\alpha_1^2 = 0.77$, $\sigma_\eta^1 = 0.21$, and $\sigma_\eta^2 = 0.25$.

The DGP2 reveals a bias in the productivity evolution parameters when cluster heterogeneity is present but not controlled for (see the Uni. GMM and the 1-comp. LIML). The Det. GMM and the 2-comp. LIML accurately control for this heterogeneity, the former more efficiently than the latter. Note that the Det. GMM does not accurately identify differences in variance across components, which is essential for valid inference. The bias in the Markov process parameters translates to a strong bias in the average cluster productivity premium. The bias of the average share-weighted productivity growth is much smaller in magnitude. In the DGP3 where cluster affiliation is known to the researcher, the 2-comp. LIML gains efficiency and obtains results almost identical to the Det. GMM. This occurs even though the cluster premia are calculated relying on model-identified rather than the imposed cluster affiliation for the 2-comp. LIML estimator.

Finally, when the analysis is based on faulty cluster affiliations (DGP4), the Det. GMM yields biased estimates. Only the 2-comp. LIML remains relatively robust. This behavior can be ascribed to the identification of cluster affiliation, which relies on the information available in the initial conditions, the evolution of productivity and the cluster affiliation indicator.

4 Application to firm-level data

Having established the performance of our estimator through MC simulations, we carry out an empirical application of the proposed estimator using balance sheet data from the Central Balance Sheet Office, VAT returns for revenue and intermediate input information, and firm-level information on employment from National Social Security Office for Belgian manufacturing firms over the period 2008-2018.

We retain a set of active firms that report output, capital stock at the beginning of the year, number of employees in full-time equivalents (FTE), and material costs.[26] This database is combined with a rich set of firm-level characteristics considered relevant for productivity growth in the literature, including firm age, industry affiliation and whether or not the firm engages in export, import and FDI activities. The data is obtained from the Belgian Balance Sheet Transaction Trade Dataset and a Belgian survey on FDI.[27] All monetary variables are deflated using the appropriate industry-level deflators constructed from national accounts.

We estimate separate production functions for five NACE Rev.2 industries—printing and reproduction of recorded media (18), manufacture of rubber and plastic products (22), manufacture of fabricated metal products, except machinery and equipment (25), manufacture of machinery and equipment n.e.c. (28) and manufacture of furniture (31)—as well as an aggregate production function for the entire manufacturing sector. These are the five largest industries in our sample

[26] We clean the data simultaneously with regards to (i) levels, (ii) ratios, and (iii) ratio growth rates to prevent the analysis from being influenced by outliers and noise. In (i), we limit the sample to observations with more than one FTE, to industry-deflated sales, materials and capital to values larger than €1,000, and export-sales and import-sales ratios up to the value of one and drop from the sample firms in industry NACE Rev. 19 (coke and refined petroleum products). In (ii), we remove the lowest and highest percentiles of the log of the labor-, capital- and materials-output ratio within NACE Rev.2 industries. In (iii), we remove observations with absolute growth rates of these ratios larger than 1000% and only retain firms that observed at least two years in a row.

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

expected to have a stationary productivity growth process.[28] We parametrize the production function $f(\cdot;\boldsymbol{\beta})$ assuming both a gross-output (Gandhi et al., 2020) and value-added (Ackerberg et al., 2015) production function under both a Cobb-Douglas and Translog specification. These production functions are estimated using either a GMM estimation approach without allowing for unobserved heterogeneity in a linear Markov process $g(\omega_{bt-1}, \boldsymbol{\alpha})$ or using the proposed LIML with increasing heterogeneity in a linear Markov process $g^s(\omega_{bt-1}, \boldsymbol{\alpha}^s)$ (with the total number of clusters S limited to $S = 10$). For space considerations and conciseness, we discuss here the estimation results for a value-added Translog production function of a specific sector (i.e., industry 22) and refer the reader to Section 5 for a complete discussion of the estimation results for all remaining specifications and industries.

4.1 Production function estimates

Table 3 presents the production function estimates required to identify productivity. The average output elasticities and returns to scale (RTS) shown in the table's first three rows indicate small, though not statistically significant, differences between the GMM and 1-comp. LIML. Most likely, these differences are linked to differences in efficiency. Specifically, the instruments for (LI)ML are constructed optimally using the nonlinear model specification $E\left[\frac{\partial \phi_{bt}}{\partial \boldsymbol{\beta}}\eta_{bt}\right] = 0$. In contrast, the GMM typically relies on factor input and output levels to specify moment conditions (Ackerberg et al., 2015; Gandhi et al., 2020).[29]

Interestingly, the production function estimates are robust to the relaxation of the homogeneity assumption concerning the productivity growth process. Comparing the output elasticities across models with increasing heterogeneity in the productivity process (1-comp. LIML up to 7-comp. LIML), it can be observed that point estimates are not identical as the number of clusters increases. They do not, however, differ significantly statistically from the 1-comp. LIML estimates. We demonstrate in Appendix E that this robustness is not specific to the methodology used in this paper.

As the number of clusters increases, we find both a minor influence on the production function coefficients and no significant effects on the shape of the productivity distribution. In particular, we report the standard deviation of the productivity estimates in the fourth row of Table 3 and the productivity ratios for firms at various percentiles of the distribution in the three subsequent rows. We observe that the ratios do not change significantly as the number of clusters increases.[30]

4.2 Latent heterogeneity in productivity

The robustness of the production function coefficients to relaxation of the homogeneity assumption concerning the productivity growth process does not imply a lack of heterogeneity in pro-

[28] See Table A.2 in the Appendix for summary statistics of this sector and industries.

[29] In this particular case, the GMM assigns relatively more importance to larger firms (in terms of input use) while the LIML assigns more weight to the fast-growing firms (in terms of input use). See also Hsiao et al. (2002) for a discussion of the difference in efficiency between GMM and ML in a dynamic panel setting.

[30] The limited variation in the shape of the productivity distribution can also be observed visually in Figures A.1 and A.2 in the Appendix where we plot the productivity densities for an increasing number of clusters.

Table 3: Estimation results based on an application with firm-level data

	GMM	LIML						
Description		1-comp.	2-comp.	3-comp.	4-comp.	5-comp.	6-comp.	7-comp.
Capital	0.132	0.121	0.126	0.127	0.126	0.127	0.126	0.126
	(0.013)	(0.018)	(0.018)	(0.016)	(0.018)	(0.018)	(0.020)	(0.018)
Labor	0.875	0.859	0.856	0.861	0.863	0.852	0.855	0.852
	(0.016)	(0.022)	(0.017)	(0.021)	(0.024)	(0.026)	(0.026)	(0.024)
RTS	1.007	0.979	0.982	0.988	0.990	0.979	0.981	0.978
	(0.011)	(0.016)	(0.016)	(0.018)	(0.021)	(0.022)	(0.024)	(0.021)
Std. Dev.	0.166	0.150	0.147	0.148	0.148	0.147	0.147	0.146
	(0.017)	(0.016)	(0.014)	(0.018)	(0.014)	(0.016)	(0.017)	(0.015)
75/25 ratio	1.014	1.013	1.013	1.013	1.013	1.013	1.013	1.013
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
95/5 ratio	1.029	1.027	1.027	1.027	1.027	1.027	1.027	1.027
	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
90/10 ratio	1.039	1.034	1.035	1.035	1.035	1.034	1.035	1.035
	(0.005)	(0.005)	(0.004)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
No. parameters	7	20	37	54	71	88	105	122
NLL		-6791	-8640	-9013	-9212	-9384	-9505	-9567
BIC		-13419	-16976	-17585	-17844	-18048	-18150	-18135
ICLbic		-13419	-16931	-17471	-17702	-17865	-17957	-17940

Notes: The first three rows display the average labor elasticities, capital elasticities, and returns to scale across firms. The fourth row displays the standard deviation of the productivity estimates. The next three rows report ratios of productivity for firms at various percentiles of the productivity distribution. Standard errors displayed between brackets are obtained using the wild bootstrap clustered at the firm level with 49 replications. No. of parameters refers to the number of parameters in the second stage of the estimation procedure. NLL stands for negative log-likelihood, BIC for Bayesian information criterion, and ICLbic for integrated complete-data likelihood Bayesian information criterion. Estimates are obtained using a panel of 626 firms and 4,399 observations in the Belgian NACE Rev. 22 industry for the period 2008-2018.

ductivity growth. The goodness-of-fit indicators reported at the bottom of Table 3 demonstrate an increasingly good model fit as the number of clusters increases, with an optimal number of six clusters as indicated by the decrease in ICLbic up to the six-comp. LIML. These six clusters are well separated, as indicated by the posterior probabilities (see Appendix A, Figure A.3).

The cluster-specific coefficients for the evolution of productivity are displayed in Table 4. We identify firm-cluster affiliation by choosing the cluster with the maximal posterior cluster affiliation probability per firm (see equation (18)). We observe heterogeneity in both the constant (α_0^s) and auto-regressive parameters (α_1^s) of the productivity process across clusters, leading to a cluster hierarchy based on stationary average productivity levels (μ_ω^s). For instance, cluster 3 has a clear productivity advantage over cluster 4, with a premium of around 20%.[31]

In addition, we observe significant heterogeneity in the volatility of the distribution of unexpected shocks to productivity (σ_η) and stationary volatility (σ_ω) that associate with stationary average productivity levels. Highly volatile productivity processes, such as those of clusters 1, 3, and 6, correlate with a relatively higher average productivity level. This indicates that firms that end up on the right tail of the productivity distribution have done so through a relatively

[31] This distinction can also be visually evaluated based on the cluster-specific productivity densities displayed in Figure A.4 in Appendix A.

Table 4: Cluster-specific characterization of the productivity evolution and its stationary distribution.

Cluster N° (s)	Prop. (%)	α_0^s	α_1^s	σ_η^s	μ_ω^s	σ_ω^s	$\bar{\omega}^s$	SD_ω^s
Cluster 1	24.804	0.794	0.939	0.042	12.957	0.121	12.984	0.130
	(5.452)	(0.169)	(0.013)	(0.005)	(0.740)	(0.021)	(0.742)	(0.022)
Cluster 2	21.268	0.645	0.950	0.025	12.839	0.080	12.832	0.071
	(6.477)	(0.492)	(0.038)	(0.006)	(0.743)	(0.035)	(0.741)	(0.029)
Cluster 3	20.988	1.064	0.918	0.071	13.000	0.179	13.024	0.159
	(3.643)	(0.209)	(0.016)	(0.010)	(0.741)	(0.020)	(0.741)	(0.017)
Cluster 4	18.238	0.915	0.929	0.018	12.803	0.047	12.816	0.066
	(1.650)	(0.383)	(0.030)	(0.004)	(0.740)	(0.018)	(0.741)	(0.021)
Cluster 5	9.147	2.405	0.813	0.038	12.838	0.065	12.826	0.060
	(2.399)	(0.625)	(0.051)	(0.008)	(0.742)	(0.018)	(0.743)	(0.018)
Cluster 6	5.556	3.949	0.696	0.144	12.993	0.200	13.018	0.193
	(0.517)	(0.428)	(0.027)	(0.024)	(0.749)	(0.036)	(0.748)	(0.036)

Notes: Prop. stands for the percentage of firms affiliated with each cluster. Standard errors displayed between brackets are obtained from a clustered wild bootstrap with 49 replications. Estimates are obtained using a panel of 626 firms and 4,399 observations in the Belgian NACE Rev. 22 industry for 2008-2018. $\mu_\omega^s = \frac{\alpha_0^s}{1-\alpha_1^s}$, $\sigma_\omega^s = \sqrt{\frac{(\sigma_\eta^s)^2}{1-(\alpha_1^s)^2}}$, $\bar{\omega}^s = \sum_{b=1}^B \sum_{t=1}^T \frac{z_b^s \omega_{bt}}{T \sum_{b=1}^B z_b^s}$, $SD_\omega^s = \sqrt{\frac{1}{T\sum_{b=1}^B z_b^s} \sum_{b=1}^B \sum_{t=1}^T z_b^s(\omega_{bt} - \bar{\omega}^s)^2}$.

volatile productivity growth process. However, high volatility in productivity does not mean that this volatility is equally persistent, as can be deduced from the auto-regressive parameters (α_1^s). Using an impulse response function in Figure 1, we demonstrate that an unexpected shock to productivity has a relatively less sizable long-lasting influence in a cluster with a volatile productivity growth process compared to one with a relatively more stable growth process.

4.3 Characterizing latent heterogeneity in productivity

Thus far, our analysis has relied on the minimal information required to estimate a production function, such as factor input and output information, which is the setting most commonly available to researchers. However, additional information is available to us regarding the age and internationalization status of firms, i.e. export, import, and/or FDI activity. We use this additional information on firm characteristics to highlight the strength of the proposed estimator and correlate it with the productivity of firm clusters.

The estimation results reported above rely on a base specification of cluster probabilities, conditioning only on initial capital, labor, and productivity (see equation (14)). This specification is derived under the assumption that the initial conditions contain sufficient information to identify cluster affiliation. If this assumption fails to hold, augmenting the base specification with additional, economically relevant (see Section 2) firm-level characteristics is necessary to help improve the identification of cluster affiliation. To test this hypothesis, we augment equa-

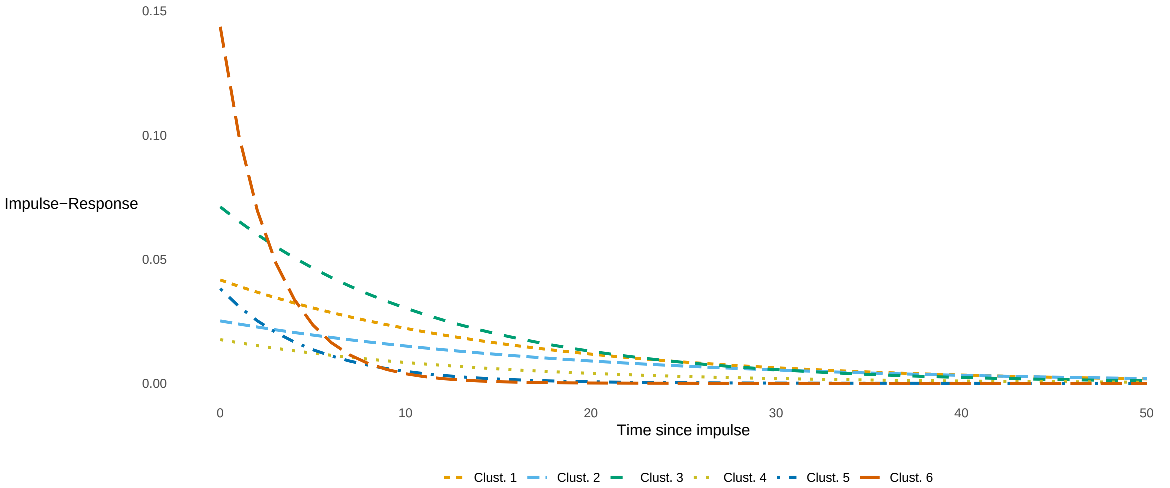

Figure 1: Impulse-response function of the cluster-specific productivity process.
Note: 50-year response function of the evolution of productivity of each cluster s after a one standard deviation unexpected productivity impulse: $IRF^s = \sum_{t=0}^{T=50} \sigma^s_\eta (\alpha^s_1)^t$. Cluster affiliation is determined as the maximal posterior cluster affiliation probability. The results shown are for estimates obtained using a panel of 626 firms and 4,399 observations in the Belgian NACE Rev. 22 industry for the period 2008-2018.

tion (14) to the following multinomial logistic specification:

$$ln\frac{Pr(z^s_b|k_{b0},l_{b0},\omega_{b0},\boldsymbol{e}_b;\boldsymbol{\gamma}^s)}{Pr(z^1_b|k_{b0},l_{b0},\omega_{b0},\boldsymbol{e}_b;\boldsymbol{\gamma}^1)} = \gamma^s_0 + \gamma^s_1 k_{b0} + \gamma^s_2 l_{b0} + \gamma^s_3 \omega_{b0} + \gamma^s_4 age_{b0} + \gamma^s_5 EXP_b + \gamma^s_6 IMP_b + \gamma^s_7 FDI_b, \quad \forall s = 2,\dots,S \tag{29}$$

where cluster probabilities are specified conditional on initial capital, labor, and productivity as well as additional firm characteristics represented in the vector $\boldsymbol{e}_b = \{age_b, EXP_b, IMP_b, FDI_b\}$, such as initial age ($age_b$), and indicators of export (EXP_b), import (IMP_b) and FDI activity (FDI_b) over the sample period.[32] Furthermore, we specify a version of (29) without initial capital, labor, and productivity. If the considered firm-level characteristics contain sufficient information to group firms into clusters, we expect this specification to perform as well as our base specification.

We rely on the two augmented specifications discussed above to re-estimate the production function. The resulting log-likelihood, BIC, and ICLbic are reported in Table 5. First, we focus on the differences between the base and augmented specifications and conclude that the former is preferred. The increase in log-likelihood obtained by the augmented specification is insufficient to warrant the increase in the number of parameters, as indicated by the smaller BIC and ICLbic indicators in absolute value relative to the base specification. The stability of the base specification to alternative specifications of cluster probabilities is in line with our Monte Carlo results and speaks to the ability of the estimator to identify firm clusters without additional information. Furthermore, this stability implies a substantial correlation between latent heterogeneity and initial conditions. Specifically, additional firm-level characteristics

[32] We classify firms as inactive or active depending on whether they are inactive over the entire sample period or active at least one point in time during the sample period. Inactive firms are chosen as the reference group.

appear to have limited explanatory power once initial conditions are controlled for.

Table 5: Goodness-of-fit indicators for estimation with varying concomitant specifications.

Specification	Log-likelihood	BIC	ICLbic
Base specification	9,504.57	-18,150.39	-17,956.60
Additional concomitants	9,515.91	-18,009.49	-17,819.55
Without initial capital and labor	9,291.26	-17,846.43	-17,591.01

Notes: The base specification refers to equation (14), the augmented specification refers to equation (29), and the specification without initial capital and labor refers to equation (29). BIC stands for Bayesian information criterion, and ICLbic for integrated complete-data likelihood Bayesian information criterion. Estimates are obtained from a value-added Translog production function with endogenous labor using a panel of 626 firms and 4,399 observations in the Belgian NACE Rev. 22 industry for 2008-2018.

To demonstrate how instrumental initial conditions perform in identifying cluster affiliation, we evaluate the model fit for the augmented specification without initial conditions. Despite the larger number of parameters compared to the base specification, the log-likelihood is smaller for this augmented specification without initial conditions, and the BIC reaffirms its superior performance. Therefore, even when firm-level information regarding age and the internationalization status of a firm is available, a significant share of the heterogeneity in productivity remains latent and cannot be accounted for using existing methods in the literature.

A closer analysis of the connection between firm characteristics and cluster affiliation can be obtained from the summary statistics across firm clusters provided in Table 6. We can deduce that initial productivity is strongly related to the stationary productivity levels of the respective clusters. The relatively low-productivity clusters (clusters 3, 4, and 5) are determined by low initial productivity, and vice versa for the relatively high-productivity clusters (clusters 1, 2, and 6). This is in line with Sterk et al. (2021), who find that initial conditions strongly determine heterogeneity in firm size.

Table 6: Average cluster characteristics

	Overall	**Clust. 1**	**Clust. 2**	**Clust. 3**	**Clust. 4**	**Clust. 5**	**Clust. 6**
Cluster prop. (%)	100.00	24.80	21.27	20.99	18.24	9.15	5.56
log(initial output)	15.18	16.16	14.66	14.84	16.04	13.55	14.91
log(inital capital)	13.30	13.99	12.92	12.88	14.17	12.35	12.91
log(initial labor)	2.78	3.48	2.66	2.07	3.93	1.67	2.14
log(initial productivity)	12.93	12.99	12.82	13.05	12.81	12.81	13.04
Inital age	24.92	27.10	25.73	21.37	29.99	19.40	23.00
Exporter prop. (%)	65.57	80.74	52.99	60.00	81.52	45.00	62.75
Importer prop. (%)	80.87	93.33	71.79	84.17	90.22	56.67	72.55
FDI prop. (%)	10.26	15.56	6.84	3.33	21.74	1.67	9.80

Notes: Cluster proportions (prop.) refer to the size of the respective clusters, where the maximal posterior cluster affiliation probability determines cluster affiliation. The results are calculated based on estimates obtained using a panel of 626 firms and 4,399 observations in the Belgian NACE Rev. 22 industry for 2008-2018.

Firm age, then, seems to be associated with the persistence of the productivity growth process. The clusters with relatively less-persistent growth processes (clusters 3, 5, and 6) contain, on average, younger firms than those with relatively persistent growth (clusters 1, 2, and 4) contain older firms on average. In addition, we observe that clusters 1 and 4—with a relatively persistent productivity growth process—associate positively with firm size in terms of initial output, capital, labor, and internationalization status, i.e., export, import, and FDI activity. Interestingly, there is a relatively large probability of importers belonging to cluster 3, which is relatively more volatile. Since the productivity estimates combine efficiency and demand drivers, we do not engage in a more detailed analysis of the anatomy of these heterogeneous effects and instead focus on understanding their economic relevance.

4.4 The impact of latent heterogeneity on exporter productivity

An intriguing observation from Table 6 is that the internationalization status of firms is associated with multiple clusters. In particular, it appears that low-productivity firms that are active in export, import, and/or FDI belong primarily to cluster 4, while higher-productivity firms with an internationalized status belong primarily to cluster 1. This observation points to heterogeneity in productivity beyond what can be captured by a simple dummy variable; a common strategy relied on in the literature (see Section 2). Lileeva and Trefler (2010) similarly document heterogeneity in the link between exporter status and the evolution of productivity.

We evaluate the economic importance of this heterogeneity by calculating the export premium (namely, the average productivity advantage of exporting over non-exporting firms in percentage terms), its evolution over time, and its contribution to aggregate productivity growth. We do this for different estimators and specifications of heterogeneity in productivity. This exercise has two purposes. First, it allows us to empirically demonstrate the robustness of the proposed methodology to the inclusion of additional information. Second, we demonstrate the importance of accounting for latent heterogeneity in productivity when comparing groups of firms that differ in specific firm-level characteristics, such as exporter status. Exporter performance has attracted the attention of multiple researchers and policymakers over the past years (see, for instance, Bernard and Bradford Jensen, 1999; Baldwin and Gu, 2003; Bernard et al., 2007; De Loecker, 2013; Garcia-Marin and Voigtländer, 2019; Gandhi et al., 2020).

We start by specifying aggregate productivity as the revenue-share weighted sum of firm-level productivities. A group of exporters (EXP) and a group of non-exporters ($NONEXP$) contribute to this aggregate productivity. Groups are indicated by $g = 1, \dots, G$, with $G = 2$ here. Aggregate productivity, then, can be decomposed into the sum of group-specific average productivities $\bar{\omega}_t^g = \sum_{b \in g} \frac{\omega_{bt}}{\sum \mathbb{I}(b \in g)}$ and within-group and between-group revenue share-productivity covariance terms, similar to Collard-Wexler and De Loecker (2015):

$$\begin{aligned}
\Omega_t &= \sum_{b=1}^{B} share_{bt}\omega_{bt} \\
&= \frac{1}{G} \sum_{g=EXP,NONEXP} \Bigg[\bar{\omega}_t^g + \underbrace{\sum_{b=1}^{B} \left(share_{bt} - \overline{share}_t^g \right) (\omega_{bt} - \bar{\omega}_t^g)}_{\text{Within-group covariance}} \\
&\qquad + \underbrace{(share_t^g - \frac{1}{G})(\Omega_t^g - \bar{\Omega}_t)}_{\text{Between-group covariance}} \Bigg],
\end{aligned} \tag{30}$$

where $share_{bt} = \frac{y_{bt}}{\sum_{b=1}^{B} y_{bt}}$, $share_t^g = \frac{\sum_{b \in g} y_{bt}}{\sum_{b=1}^{B} y_{bt}}$, $\overline{share}_t^g = \frac{1}{G} \sum_g share_t^g$, $\Omega_t^g = \sum_{b \in g} share_{bt} \omega_{bt}$, and $\bar{\Omega}_t = \frac{1}{G} \sum_g \Omega_t^g$. The within-group revenue share-productivity covariance term captures the covariance between the revenue share and productivity within each group of exporters and non-exporters. A positive within-group covariance indicates that more productive firms also hold larger market shares. The between-group revenue share-productivity covariance term captures the covariance of the aggregate revenue share and productivity between the groups of exporters and non-exporters. A positive between-group covariance indicates that the more productive groups also hold more market share.

We calculate this decomposition of aggregate productivity for different productivity indices obtained from different estimation methodologies and different specifications of heterogeneity in productivity.[33] Specifically, we estimate productivity using the GMM and LIML identification strategies with (i) a *base* specification: $\omega_{bt} = \alpha_0 + \alpha_1 \omega_{bt-1} + \eta_{bt}$, (ii) a *deterministic* control for exporter status, $\omega_{bt} = \alpha_0 + \alpha_1 \omega_{bt-1} + \alpha_2 EXP_b + \alpha_3 \omega_{bt-1} EXP_b + \eta_{bt}$, and (iii) a more *exhaustive* set of controls for heterogeneity in productivity:

$$\begin{aligned}\omega_{bt} = \alpha_0 + \alpha_1 \omega_{bt-1} + \alpha_3 age_{b0} + \alpha_4 Exp_b + \alpha_5 \omega_{bt-1} Exp_b \\ + \alpha_6 Imp_b + \alpha_7 \omega_{bt-1} Imp_b \\ + \alpha_8 FDI_b + \alpha_9 \omega_{bt-1} FDI_b + \eta_{bt}.\end{aligned} \tag{31}$$

Similarly, we obtain productivity from the finite mixture LIML identification strategy with the optimal number of six clusters and (i) the *base* specification for cluster affiliation (14), (ii) the base specification for cluster affiliation augmented with a *deterministic* control for internationalization status using a dummy indicator, and (iii) an *exhaustive* control for heterogeneity in the specification for cluster affiliation (29).

Figure 2 displays the evolution of the obtained aggregate productivities and their decomposition across estimation methodologies and specifications. Focusing on aggregate productivity (the left column), we observe that the evolution over time is very similar across estimation methodologies and specifications. This behavior can be attributed to the robustness of the production function and productivity estimates to the homogeneity assumption of the evolution of productivity, as reported above. This behavior is also in line with the MC analysis (see subsection 3.5), based on which we expect a strong bias in the productivity premium, but a smaller bias in average share-weighted productivity growth.

Exploring the decomposition of this aggregate productivity, we observe differences depending on the estimation methodology and the specification of heterogeneity in the Markov process, for both the GMM and LIML estimation methodologies. This contrasts with the robustness of the finite mixture LIML across specifications. For instance, the export premium—the difference between the average productivity of exporters (dashed line) and non-exporters (continuous line) in the second column of Figure 2— evolves from 1.97%, for the base specification, to 3.16% for the deterministic and 2.16% for the exhaustive specification of heterogeneity for

[33] For each estimation methodology and specification, we normalize aggregate productivity relative to share-weighted aggregate productivity in the initial year (Ω_0) (Aw et al., 2001).

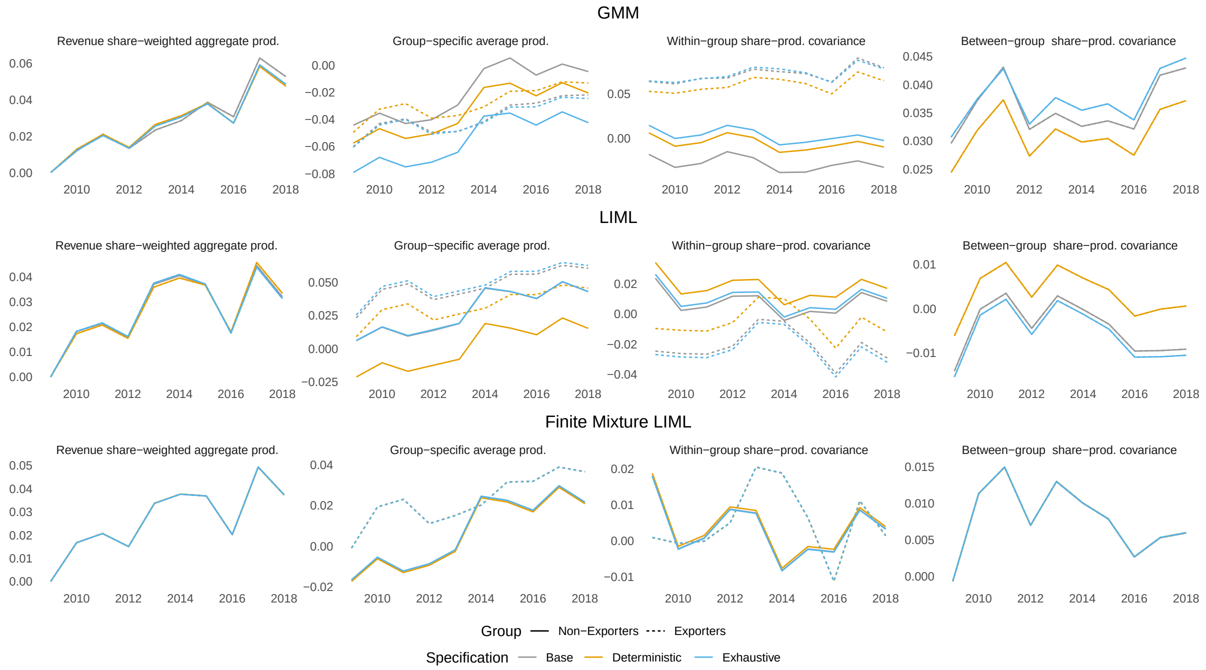

Figure 2: Evolution of aggregate productivity and its decomposition for exporting- and non-exporting firms.

Notes: GMM, LIML, and finite mixture LIML refer to the productivity estimation methodologies, while Base, Deterministic, and Exhaustive refer to the specification of heterogeneity within these methodologies, i.e., see equations (31) and (29).

the LIML methodology (see Appendix A.3). Notably, the observed differences in export premia between heterogeneity specifications are relatively constant over time. In comparison, the export premium is approximately 1.6% for all three specifications of the finite mixture LIML.

As a result of the observed variability in export premia, the within-group and between-group covariance terms for the GMM and LIML estimation methodologies are dependent on the heterogeneity specification and attain negative values for some specifications. The finite mixture LIML methodology, on the other hand, reports a slightly positive and robust within- and between-group covariance term, meaning more productive exporters have, on average, a greater market share.

Overall, latent heterogeneity does not strongly affect the evolution of measured aggregate productivity or of the export premium over time. It does, however, affect export premium levels, the separate components of the aggregate productivity decomposition, and, subsequently, conclusions drawn regarding misallocation issues across firms. As such, correctly controlling for latent heterogeneity in productivity is of interest to any applied researcher or policymaker interested in productivity premia between groups of firms and their respective contribution to aggregate productivity growth as a driver of economic growth and welfare.

5 Robustness

This paper reports the estimation results for a value-added Translog production function for NACE Rev.2 industry 22. We demonstrate in Appendix D that the reported results are robust to the estimation methodology and industry selection. We evaluate the results for four alternative estimation methodologies, assuming both gross output and value-added under both a Cobb-Douglas and Translog specification, for all manufacturing industries considered. The proposed method delivers economically sensible production function estimates in all cases and confirms the results presented.

It could be of concern that our results are specific to the Belgian firm-level dataset. Therefore, we also apply the analysis to the Chilean firm-level dataset used by Gandhi et al. (2020). These results reaffirm the findings obtained with the Belgian firm-level dataset where we find little evidence of a significant omitted variable bias, but strong evidence favoring multiple clusters in the productivity evolution (see Appendix D).

6 Conclusion

This paper proposes a general extension of state-of-the-art production function estimation procedures to control for and identify latent heterogeneity in the evolution of productivity. We demonstrate the applicability of this methodology by means of a Monte Carlo simulation and an application to Belgian firm-level data. We find strong evidence of latent heterogeneity in the evolution of productivity. This unobserved heterogeneity is associated with the initial conditions of a firm, especially the starting level of productivity. The uncovered importance of ex-ante heterogeneity relative to ex-post shocks is in line with earlier literature and becomes relevant for understanding the macroeconomic effects of firm-level frictions.

Additional explanatory variables expected to capture differences in the evolution of productivity, such as the export, import, and FDI status of a firm, are associated with multiple productivity clusters obtained from the proposed method. This indicates heterogeneity in productivity beyond what is captured by the observed firm-level characteristics. As a result, current productivity estimation methodologies depend on such additional firm-level information which remains notoriously unavailable, especially for characteristics that are hard to quantify, such as intangible capital or managerial capacity. The proposed methodology, on the other hand, maintains its performance irrespective of the presence of this type of supplementary information.

Building on the newly developed productivity estimation strategy, one can systematically search for the main determinants of productivity growth, which is accurately identified along with its underlying clusters. Obtaining such insight is based on notions of similarity and dissimilarity between firms and groups of firms. Firms in the same cluster share the same growth process and are thus "similar", while heterogeneity allows "dissimilar" firms to grow at a different pace across clusters. The advantage of this approach is its flexibility to allow for and identify an unobserved firm cluster structure. Conversely, current methods work with a predefined cluster of firms—such as industry-specific productivity growth processes—and aim to find within-cluster determinants of productivity growth. To that end, the proposed approach allows the data to determine firm clusters and identify the *between-cluster* determinants of productivity growth, i.e. the firm-level characteristics that drive cluster affiliation.

As such, the methodology proposed in this paper opens up exciting new avenues for research. It is of relevance to every applied researcher interested in accurately recovering the effects of a firm-specific event (e.g. engagement in export activity) on the evolution of firm-level productivity by accounting for unobserved heterogeneity in productivity. However, while the proposed methodology allows us to correctly identify unobserved heterogeneity in productivity, further work is needed to fully understand the drivers of this heterogeneity.

References

Ackerberg, D., C. Lanier Benkard, S. Berry, and A. Pakes (2007). Chapter 63 econometric tools for analyzing market outcomes. Volume 6 of Handbook of Econometrics, pp. 4171–4276. Elsevier.

Ackerberg, D. A. (2021). Comment on "olley and pakes-style production function estimators with firm fixed effects". Oxford Bulletin of Economics and Statistics 83(3), 836–840.

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification Properties of Recent Production Function Estimators. Econometrica 83(6), 2411–2451.

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

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

Arcidiacono, P., P. B. Ellickson, et al. (2011). Practical methods for estimation of dynamic discrete choice models. Annual Review of Economics 3(1), 363–394.

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

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

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. Journal of Econometrics 68(1), 29–51.

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

Aw, B. Y., X. Chen, and M. J. Roberts (2001). Firm-level evidence on productivity differentials and turnover in taiwanese manufacturing. Journal of Development Economics 66(1), 51–86.

Aw, B. Y., M. J. Roberts, and D. Y. Xu (2011). R&d investment, exporting, and productivity dynamics. American Economic Review 101(4), 1312–1344.

Baldwin, J. R. and W. Gu (2003). Export-market participation and productivity performance in canadian manufacturing. Canadian Journal of Economics/Revue canadienne d'économique 36(3), 634–657.

Bas, M., T. Mayer, and M. Thoenig (2017). From micro to macro: Demand, supply, and heterogeneity in the trade elasticity. Journal of International Economics 108, 1 – 19.

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.

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.

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

Bernard, A. B. and J. Bradford Jensen (1999). Exceptional exporter performance: cause, effect, or both? Journal of International Economics 47(1), 1–25.

Bernard, A. B., J. B. Jensen, S. J. Redding, and P. K. Schott (2007, September). Firms in international trade. Journal of Economic Perspectives 21(3), 105–130.

Bernard, A. B., S. J. Redding, and P. K. Schott (2009). Products and productivity. The Scandinavian Journal of Economics 111(4), 681–709.

Biernacki, C., G. Celeux, and G. Govaert (2000). Assessing a mixture model for clustering with the integrated completed likelihood. IEEE transactions on pattern analysis and machine intelligence 22(7), 719–725.

Bilir, L. K. and E. Morales (2020). Innovation in the global firm. Journal of Political Economy 128(4), 1566–1625.

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.

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

Blum, B. S., S. Claro, I. Horstmann, and D. A. Rivers (2021). The abcs of firm heterogeneity when firms sort into markets: The case of exporters. Technical report, Working Paper.

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 87(1), 115–143.

Brandt, L., J. Van Biesebroeck, L. Wang, and Y. Zhang (2017, September). Wto accession and performance of chinese manufacturing firms. American Economic Review 107(9), 2784–2820.

Caliendo, L., G. Mion, L. D. Opromolla, and E. Rossi-Hansberg (2020). Productivity and organization in portuguese firms. Journal of Political Economy 128(11), 4211–4257.

Celeux, G., S. Frühwirth-Schnatter, and C. P. Robert (2018). Model selection for mixture models—perspectives and strategies. In Handbook of mixture analysis, pp. 121–160. CRC Press.

Collard-Wexler, A. and J. De Loecker (2015, January). Reallocation and technology: Evidence from the us steel industry. American Economic Review 105(1), 131–71.

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.

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

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

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.

De Loecker, J. and P. K. Goldberg (2014). Firm performance in a global market. Annual Review of Economics 6(1), 201–227.

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

De Loecker, J. and C. Syverson (2021). An industrial organization perspective on productivity. In K. Ho, A. Hortaçsu, and A. Lizzeri (Eds.), Handbook of Industrial Organization, Volume 4, Volume 4 of Handbook of Industrial Organization, pp. 141–223. Elsevier.

Dewitte, R., M. Dumont, B. Merlevede, G. Rayp, and M. Verschelde (2020). Firm-heterogeneous biased technological change: A nonparametric approach under endogeneity. European Journal of Operational Research 283(3), 1172–1182.

Dewitte, R., M. Dumont, G. Rayp, and P. Willemé (2022). Unobserved heterogeneity in the productivity distribution and gains from trade. Canadian Journal of Economics/Revue canadienne d'économique 55(3), 1566–1597.

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

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

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

Frühwirth-Schnatter, S., C. Pamminger, A. Weber, and R. Winter-Ebmer (2012). Labor market entry and earnings dynamics: Bayesian inference using mixtures-of-experts markov chain clustering. Journal of Applied Econometrics 27(7), 1116–1137.

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.

Garcia-Marin, A. and N. Voigtländer (2019). Exporting and plant-level efficiency gains: It's in the measure. Journal of Political Economy 127(4), 1777–1825.

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

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

Harrigan, J., A. Reshef, and F. Toubal (2018, November). Techies, trade, and skill-biased productivity. Working Paper 25295, National Bureau of Economic Research.

Haskel, J. and S. Westlake (2017). Capitalism without capital. In Capitalism without Capital. Princeton University Press.

Hsiao, C., M. Hashem Pesaran, and A. Kamil Tahmiscioglu (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. Journal of Econometrics 109(1), 107–150.

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.

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.

Kasahara, H., M. Nishida, and M. Suzuki (2017). Decomposition of aggregate productivity growth with unobserved heterogeneity. Technical Report 17-E-083.

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.

Kasahara, H., P. Schrimpf, and M. Suzuki (2015). Identification and estimation of production function with unobserved heterogeneity. Technical report.

Klette, T. J. and Z. Griliches (1996). The Inconsistency of Common Scale Estimators When Output Prices Are Unobserved and Endogenous. Journal of Applied Econometrics 11(4), 343–61.

Konings, J. and S. Vanormelingen (2015, 05). The Impact of Training on Productivity and Wages: Firm-Level Evidence. The Review of Economics and Statistics 97(2), 485–497.

Lee, Y., A. Stoyanov, and N. Zubanov (2019). Olley and pakes-style production function estimators with firm fixed effects. Oxford Bulletin of Economics and Statistics 81(1), 79–97.

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

Li, D., L. Simar, and V. Zelenyuk (2016). Generalized nonparametric smoothing with mixed discrete and continuous data. Computational Statistics & Data Analysis 100, 424–444.

Lileeva, A. and D. Trefler (2010, 08). Improved Access to Foreign Markets Raises Plant-level Productivity…For Some Plants*. The Quarterly Journal of Economics 125(3), 1051–1099.

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

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

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

Merlevede, B. and A. Theodorakopoulos (2020). Intangibles within firm boundaries. Oxford Martin Working Paper Series on Economic and Technological Change 6, Oxfort Martin School.

Merlevede, B. and A. Theodorakopoulos (2021). Productivity effects of internationalisation through the domestic supply chain. Journal of Applied Econometrics 36(6), 808–832.

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

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.

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

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

Rubens, M. (2020). Managerial knowledge and technology choice: evidence from us mining schools. FEB Research Report Department of Economics.

Rubens, M. (2021). Market structure, oligopsony power, and productivity. Oligopsony Power, and Productivity (March 8, 2021).

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

Sterk, V., P. Sedláček, and B. Pugsley (2021, February). The nature of firm growth. American Economic Review 111(2), 547–579.

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

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

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

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

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

Wooldridge, J. M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. Journal of applied econometrics 20(1), 39–54.

Online Appendix
Identifying Unobserved Heterogeneity in Productivity

Ruben Dewitte, Catherine Fuss and Angelos Theodorakopoulos*

December, 2022

*Dewitte (corresponding author): Ghent University, Tweekerkenstraat 2, 9000 Ghent, Belgium, rubenl.dewitte@ugent.be; Fuss: National Bank of Belgium, 14 Boulevard de Berlaimont, 1000 Brussels, Belgium, catherine.fuss@nbb.be; Theodorakopoulos: Aston Business School, 295 Aston Express Way, Birmingham B4 7ER, UK, a.theodorakopoulos2@aston.ac.uk. The authors would like to thank Emmanuel Dhyne, Cédric Duprez, Rebecca Freeman, David Rivers, an anonymous referee and participants in the 2021 NAPW conference, the 2021 ITIM seminar at Ghent University, the 2022 Antwerp University seminar, the 2022 BORDERS workshop at Ghent University and the 2022 CEA conference at Carleton University. Ruben Dewitte gratefully acknowledges financial support from the Research Foundation Flanders (FWO) grant 12B8822N and the National Bank of Belgium (NBB). This paper was completed during the NBB's internship program for young researchers. The views expressed herein are those of the authors and do not necessarily reflect the views of the NBB or any other institution to which the authors are affiliated. Any errors are the authors' own.

Appendix A Additional Figures and Tables

A.1 Figures

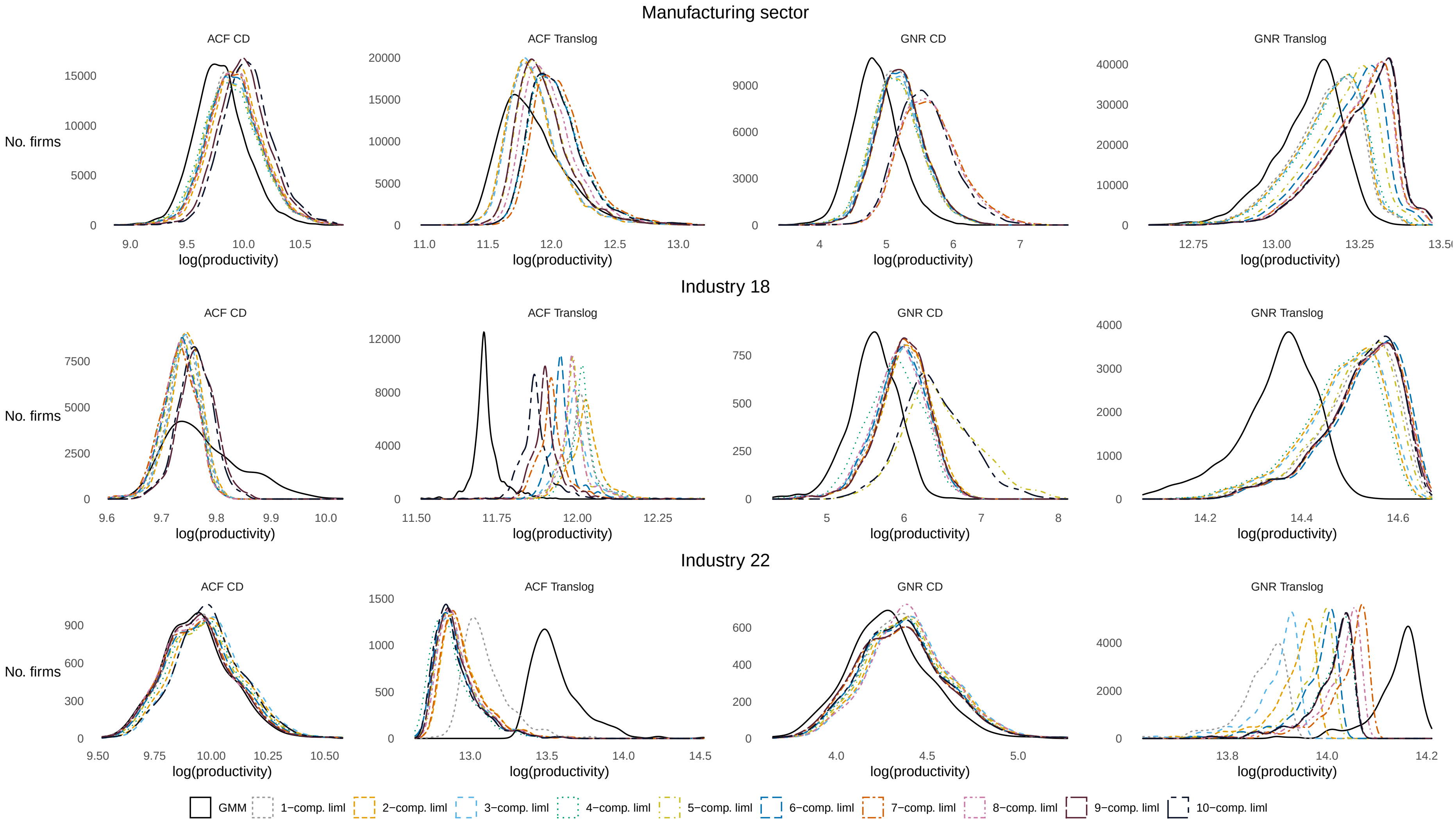

Figure A.1: Density of 1- to 10-cluster productivity (ω_{bt}) in 2013 obtained from Ackerberg et al. (2015) and Gandhi et al. (2020) methods under Cobb-Douglas and Translog production functions and endogenous labor for the entire manufacturing sector and industries 18 and 22 of the Belgian economy.

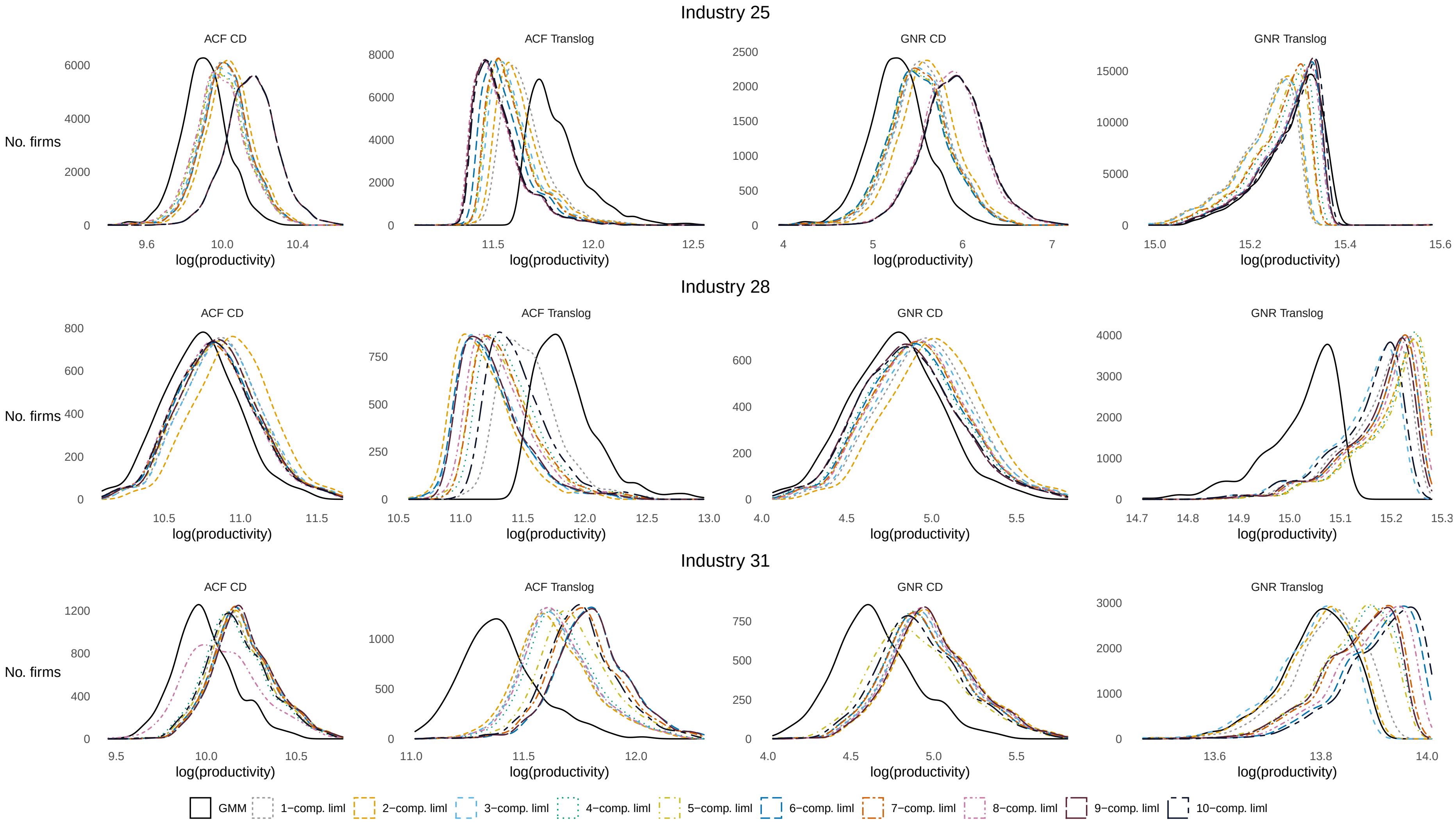

Figure A.2: Density of 1- to 10-cluster productivity (ω_{bt}) in 2013 obtained from Ackerberg et al. (2015) and Gandhi et al. (2020) methods under Cobb-Douglas and Translog production functions and endogenous labor for the entire manufacturing sector and industries 25, 28, and 31 of the Belgian economy.

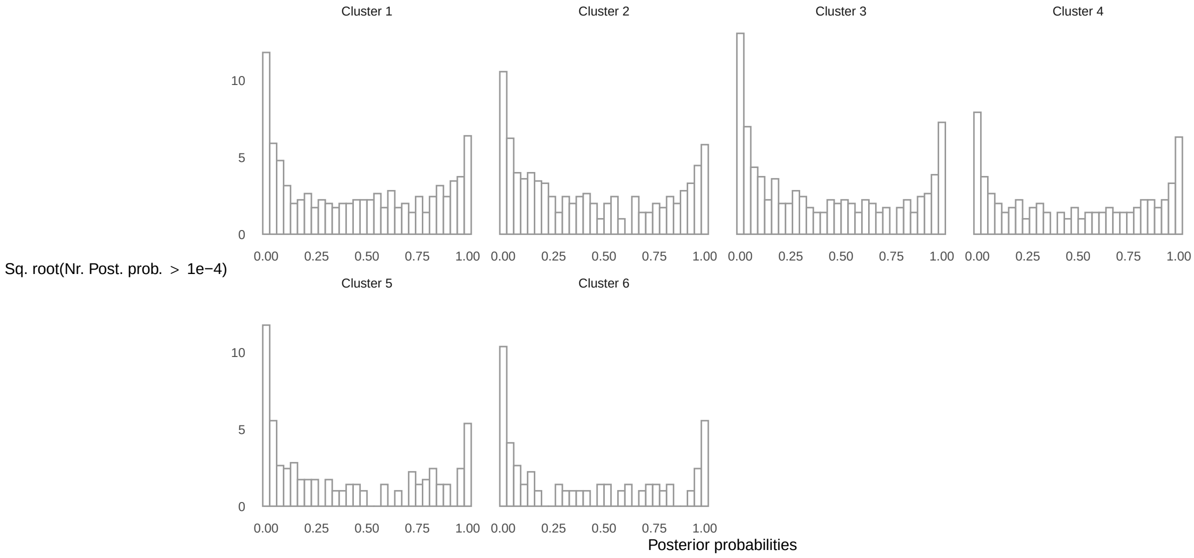

Figure A.3: Histogram of posterior probabilities for a 6-cluster (ACF) value-added Translog production function of NACE Rev. 22 estimated with LIML.

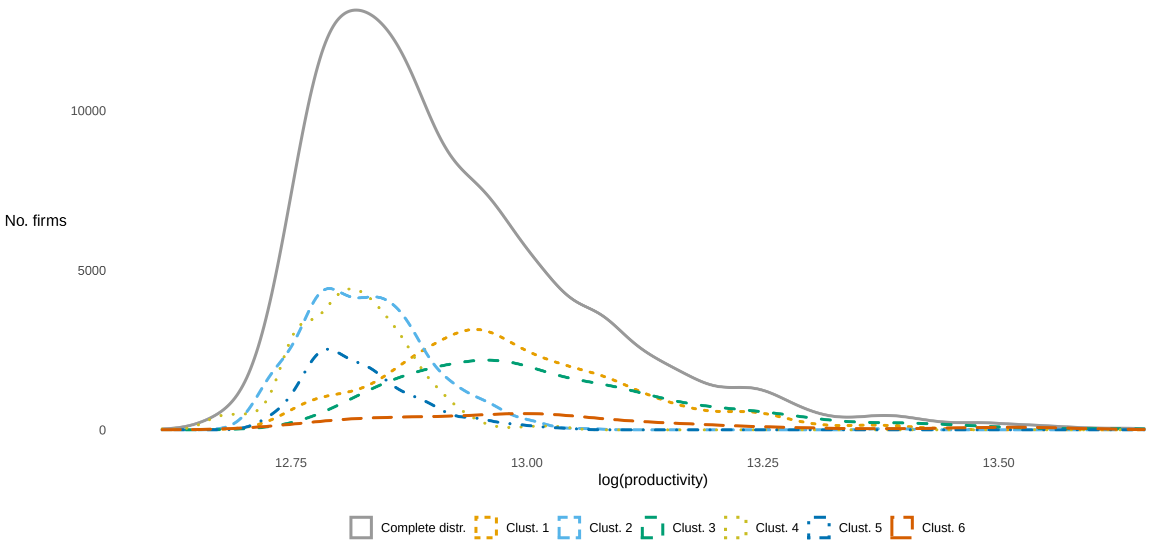

Figure A.4: Complete and cluster-specific density of productivity in 2013 obtained from a 6-cluster value-added Translog production function of NACE Rev. 22 estimated with LIML.

A.2 Tables

Table A.1: Description of mathematical symbols

Symbol	**Description**
$b = 1, \ldots, B$	Firm identifier
$t = 1, \ldots, T$	Time indicator
$s = 1, \ldots, S$	Cluster indicator
ω_{bt}	Hicks-neutral total factor productivity (TFP)
ε_{bt}	Ex-post shock to production
η_{bt}	Innovation/productivity shock
$\xi_{bt}(\cdot)$	Choice-specific i.i.d. variable that captures the affinity of a firm for a certain cluster
ζ_{bt}	Error term of the reduced-form labor equation
$\boldsymbol{\epsilon}^s \equiv \begin{bmatrix} \varepsilon_{bt} \\ \eta_{bt} \end{bmatrix}$	Cluster s-specific overarching vector of error terms
$\boldsymbol{\Sigma}^s = \begin{bmatrix} (\sigma^s_\eta)^2 & \sigma^s_{\eta,\zeta} \\ \sigma^s_{\eta,\zeta} & (\sigma^s_\zeta)^2 \end{bmatrix}$	Cluster s-specific variance-covariance matrix
σ_x	Standard deviation of variable x
$\boldsymbol{e}_{bt}$	Vector of latent firm-level drivers of productivity
$\mathcal{I}$	A firm's information set at time t
Y_{bt}	Firm-level output
$X_{bt} \in \{K_{bt}, L_{bt}, M_{bt}\}$	Generic input respectively representing capital, labor or materials
$z^i_b = \mathbb{I}_b\left(s = i\right), \forall i = 1, \ldots, S$	Firm-level cluster affiliation indicator
$\beta \in (0, 1)$	Discount factor
$\pi_t(\cdot)$	Profit function giving current-period profits as a function of the vector of state variables
ϕ_{bt}	Remaining output variation after netting out the estimates of the first stage ex-post shocks to production and, for the case of a gross-output production function, the output contribution of the flexible production factor.
$\boldsymbol{\beta}$	Vector of production function parameters
$\boldsymbol{\alpha}$	Vector of productivity process parameters
$\boldsymbol{\theta}$	Vector of parameters for the reduced-form labor equation
$\boldsymbol{\theta}^s \equiv \{\boldsymbol{\beta}, \boldsymbol{\alpha}^s, \boldsymbol{\delta}^s, \boldsymbol{\Sigma}^s\}$	Overarching vector of parameters
$\boldsymbol{\gamma}$	Vector of cluster probability parameters
$\boldsymbol{\Theta} \equiv \left\{\boldsymbol{\gamma}^1, \ldots, \boldsymbol{\gamma}^S, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^S\right\}$	Overarching vector of parameters for the main estimation equation
$\mathbb{I}(\cdot)$	Indicator function
$p(\cdot)$	Probability density function
$\varphi(\cdot)$	Standard-normal probability density function

Table A.2: Summary Statistics

Industry	Variable	# Obs.	# Firms	Min.	Q_{25}	Median	Q_{75}	Max	Mean	sd
Manufacturing sector	Log(Sales)	103170	14344	10.28	13.26	14.17	15.36	22.90	14.45	1.60
Manufacturing sector	Log(Employment)	103170	14344	0.22	1.32	2.11	3.15	8.99	2.34	1.33
Manufacturing sector	Log(Capital)	103170	14344	6.91	11.44	12.67	13.78	20.61	12.64	1.86
Manufacturing sector	Log(Materials)	103170	14344	9.11	12.67	13.71	14.99	22.60	13.97	1.73
Industry 18	Log(Sales)	7321	1109	11.14	12.99	13.59	14.51	18.62	13.84	1.21
Industry 18	Log(Employment)	7321	1109	0.22	1.01	1.66	2.53	6.24	1.87	1.10
Industry 18	Log(Capital)	7321	1109	6.92	11.25	12.47	13.49	18.25	12.37	1.70
Industry 18	Log(Materials)	7321	1109	9.91	12.46	13.12	14.14	18.54	13.36	1.31
Industry 22	Log(Sales)	4310	616	11.70	14.15	15.10	16.23	19.93	15.24	1.51
Industry 22	Log(Employment)	4310	616	0.22	1.83	2.80	3.76	7.19	2.86	1.36
Industry 22	Log(Capital)	4310	616	7.29	12.22	13.39	14.61	17.90	13.35	1.79
Industry 22	Log(Materials)	4310	616	10.51	13.72	14.76	15.96	19.67	14.86	1.61
Industry 25	Log(Sales)	21357	3197	10.93	13.33	14.07	14.90	20.51	14.19	1.23
Industry 25	Log(Employment)	21357	3197	0.22	1.39	2.08	2.88	8.21	2.18	1.12
Industry 25	Log(Capital)	21357	3197	6.91	11.38	12.56	13.48	18.19	12.41	1.60
Industry 25	Log(Materials)	21357	3197	9.29	12.72	13.58	14.48	20.20	13.67	1.36
Industry 28	Log(Sales)	5781	954	11.54	13.93	14.74	15.80	21.40	14.93	1.42
Industry 28	Log(Employment)	5781	954	0.22	1.61	2.48	3.44	8.12	2.58	1.31
Industry 28	Log(Capital)	5781	954	6.99	11.55	12.76	13.76	19.24	12.63	1.70
Industry 28	Log(Materials)	5781	954	9.76	13.39	14.25	15.38	21.24	14.41	1.51
Industry 31	Log(Sales)	5558	806	11.07	13.24	13.96	14.80	18.95	14.13	1.22
Industry 31	Log(Employment)	5558	806	0.22	1.25	2.01	2.89	5.81	2.15	1.14
Industry 31	Log(Capital)	5558	806	6.94	11.30	12.42	13.36	17.25	12.30	1.57
Industry 31	Log(Materials)	5558	806	10.14	12.68	13.47	14.42	18.77	13.65	1.32

Table A.3: Average export premia across productivity estimation methodologies and specifications

Methodology	Specification	Industry 18	Industry 22	Industry 25	Industry 28	Industry 31
GMM	Base	0.0032	-0.0179	-0.0039	0.0520	0.0108
		(0.0009)	(0.0041)	(0.0022)	(0.0062)	(0.0027)
GMM	Deterministic	0.0032	0.0061	0.0162	0.0783	0.0838
		(0.0009)	(0.0036)	(0.0020)	(0.0060)	(0.0028)
GMM	Exhaustive	0.0032	0.0160	0.0246	0.1071	0.0932
		(0.0009)	(0.0035)	(0.0020)	(0.0053)	(0.0026)
LIML	Base	-0.0153	0.0197	0.0260	0.1048	-0.0086
		(0.0008)	(0.0034)	(0.0022)	(0.0068)	(0.0035)
LIML	Deterministic	0.1053	0.0316	0.0382	0.1420	-0.0056
		(0.0014)	(0.0033)	(0.0021)	(0.0065)	(0.0034)
LIML	Exhaustive	0.1045	0.0216	0.0438	0.1296	0.0284
		(0.0014)	(0.0034)	(0.0021)	(0.0067)	(0.0029)
Finite Mixture LIML	Base	-0.0132	0.0160	0.0098	0.0467	-0.0157
		(0.0009)	(0.0034)	(0.0021)	(0.0070)	(0.0036)
Finite Mixture LIML	Deterministic	-0.0113	0.0167	0.0100	0.0472	-0.0145
		(0.0009)	(0.0034)	(0.0021)	(0.0070)	(0.0036)
Finite Mixture LIML	Exhaustive	-0.0075	0.0159	0.0095	0.0440	-0.0138
		(0.0009)	(0.0034)	(0.0021)	(0.0070)	(0.0036)

Notes: Export premia obtained from a log-linear regression with year dummies t, $\omega_{bt} = \alpha Exp_b + t + \epsilon_{bt}$, where productivity is obtained from GMM, LIML and Finite Mixture LIML estimation methodologies with a base, deterministic or exhaustive specification of heterogeneity within these methodologies, i.e. see equations (31) and (29). Standard errors between brackets are obtained from the OLS regression.

Appendix B Productivity estimation

This section describes the production function estimation techniques used in this paper. We summarize the proxy variable (Ackerberg et al., 2015) and the first-order condition (Gandhi et al., 2020) methods before advancing to our proposed Mixture LIML estimator.

B.1 Proxy variable methods

Following Ackerberg et al. (2015), the production function is assumed Leontief in (and proportional to) materials, that is,

$$Y_{bt} = min\left\{F^{kl}\left(K_{bt}, L_{bt}\right), M_{bt}\right\} e^{\omega_{bt}+\varepsilon_{bt}}, \tag{B.1}$$

Provided that the Leontief first-order condition holds, the (log) valued-added production function to be estimated is:

$$\frac{y_{bt}}{m_{bt}} = f^{kl}\left(k_{bt}, l_{bt}\right) + \omega_{bt} + \varepsilon_{bt}. \tag{B.2}$$

Materials are assumed to be a flexible factor input, while capital and labor are considered as non-flexible. As such, materials are chosen at time t based on the available information set $\mathcal{I}_{bt}$, including current productivity $\omega_{bt} \in \mathcal{I}_{bt}$. From the materials input demand, assuming scalar unobservability and strict monotonicity between material demand and productivity, it follows that materials can proxy for productivity (Ackerberg et al., 2007):

$$m_{bt} = h\left(\mathcal{I}_{bt}\right), \qquad \omega_{bt} = h^{-1}\left(m_{bt}, \mathcal{I}_{bt} \setminus \omega_{bt}\right). \tag{B.3}$$

The estimation strategy proposed by Ackerberg et al. (2015), consists of two stages. In the *first stage*, one relies on the materials as a proxy for productivity to single out the ex-post Hicks-neutral productivity shock and possible classical measurement error ε_{bt}:

$$\frac{y_{bt}}{m_{bt}} = f^{kl}\left(k_{bt}, l_{bt}\right) + h^{-1}\left(m_{bt}, k_{bt}, l_{bt}\right) + \varepsilon_{bt}. \tag{B.4}$$

The consistent estimates from this first stage estimation allow us to retrieve the non-flexible output (log value-added) variation:

$$\phi_{bt} \equiv \frac{y_{bt}}{m_{bt}} - \varepsilon_{bt} = f^{kl}\left(k_{bt}, l_{bt}\right) + \omega_{bt}. \tag{B.5}$$

Assuming that productivity evolves according to a first-order Markov process, $\omega_{bt} = g(\omega_{bt-1}) + \eta_{bt}$, this results in the *second stage* estimation equation:

$$\phi_{bt} = f^{kl}\left(k_{bt}, l_{bt}\right) + g\left(\phi_{bt-1} - f^{kl}\left(k_{bt-1}, l_{bt-1}\right)\right) + \eta_{bt}. \tag{B.6}$$

Consistent parameter estimates for the production function can be obtained building on the moment conditions following from the independence between the timing of factor input decisions and the unexpected shocks to productivity:

$$E\left[\eta_{bt}|k_{bt}, l_{bt(-1)}, \phi_{bt-1}\right] = 0. \tag{B.7}$$

We parametrize equation (B.6) with production function coefficients $\boldsymbol{\beta}$ and specify a linear first-order Markov process (see equation (5) in main text) $g(\phi_{bt-1}, l_{bt-1}, k_{bt-1}; \boldsymbol{\beta}, \boldsymbol{\alpha}) = \boldsymbol{W}_{bt-1}\boldsymbol{\alpha}$ with $\boldsymbol{W}_{bt-1} = [1, \omega_{bt-1}]$, and $\theta = \{\boldsymbol{\beta}, \boldsymbol{\alpha}\}$ such that

$$\phi_{bt} = f^{kl}\left(k_{bt}, l_{bt}; \boldsymbol{\beta}\right) + \boldsymbol{W}_{bt-1}\boldsymbol{\alpha} + \eta_{bt}. \tag{B.8}$$

This equation is linear in the Markov process parameters $\boldsymbol{\alpha}$ and non-linear in the production function parameters $\boldsymbol{\beta}$. To speed up the estimation procedure by reducing the non-linear parameter space, we iteratively search for the optimal, non-linear, production function parameters and, given the production function parameter estimates, rely on a closed-form solution for the linear Markov process parameters at each iteration.

First, we specify the optimization problem for the production function parameters, $\boldsymbol{\beta}$. With instrumental variables $\boldsymbol{Z}_{bt} = \left[k_{bt}, l_{bt(-1)}\right]$ and a weighting matrix $\left(\frac{\boldsymbol{Z}_{bt}^T\boldsymbol{Z}_{bt}}{B}\right)^{-1}$, the optimization criterion is:

$$\arg\min_{\boldsymbol{\beta}} \Lambda(\boldsymbol{\beta}) = \arg\min_{\boldsymbol{\beta}} \left(\frac{\sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{Z}_{bt}^T\eta_{bt}}{B}\right)^T \left(\frac{\sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{Z}_{bt}^T\boldsymbol{Z}_{bt}}{B}\right)^{-1} \left(\frac{\sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{Z}_{bt}^T\eta_{bt}}{B}\right), \tag{B.9}$$

with the corresponding First-Order Condition (FOC):

$$\begin{aligned}
\nabla_{\boldsymbol{\beta}}\Lambda(\theta) = 0 = -2\left(\frac{1}{B}\sum_{b=1}^{B}\sum_{t=1}^{T}(\boldsymbol{Z}_{bt})^T\nabla_{\boldsymbol{\beta}}\eta_{bt}\right)\left(\frac{(\boldsymbol{Z}_{bt})^T\boldsymbol{Z}_{bt}}{B}\right)^{-1}\left(\frac{1}{B}\sum_{b=1}^{B}\sum_{t=1}^{T}(\boldsymbol{Z_{bt}})^T\eta_{bt}\right)\\
\Leftrightarrow\\
0 = \left(\frac{1}{B}\sum_{b=1}^{B}\sum_{t=1}^{T}(\boldsymbol{Z}_{bt})^T\nabla_{\boldsymbol{\beta}}\eta_{bt}\right)\left(\frac{(\boldsymbol{Z}_{bt})\boldsymbol{Z}_{bt}}{B}\right)^{-1}\left(\frac{1}{B}\sum_{b=1}^{B}\sum_{t=1}^{T}(\boldsymbol{Z}_{bt})^T\eta_{bt}\right),
\end{aligned}$$

where $\nabla_{\boldsymbol{\beta}}(\eta_{bt}) = -\nabla_{\boldsymbol{\beta}}f^{kl}\left(k_{bt}, l_{bt}; \boldsymbol{\beta}\right) + \alpha_1\nabla_{\boldsymbol{\beta}}f^{kl}\left(k_{bt-1}, l_{bt-1}; \boldsymbol{\beta}\right) + \boldsymbol{W}_{bt-1}\nabla_{\boldsymbol{\beta}}\boldsymbol{\alpha}$.

In every iteration, we optimize for the Markov process parameters, $\boldsymbol{\alpha}$, given a value of the production function parameters, $\boldsymbol{\beta}$. The optimization criterion with weighting matrix $\left(\frac{\boldsymbol{W}_{bt}^T\boldsymbol{W}_{bt}}{B}\right)^{-1}$ is:

$$\arg\min_{\boldsymbol{\alpha}} \Lambda(\boldsymbol{\alpha}) = \arg\min_{\boldsymbol{\alpha}} \left(\frac{\sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{W}_{bt}^T\eta_{bt}(\hat{\boldsymbol{\beta}})}{B}\right)^T \left(\frac{\sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{W}_{bt}^T\boldsymbol{W}_{bt}}{B}\right)^{-1} \left(\frac{\sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{W}_{bt}^T\eta_{bt}(\hat{\boldsymbol{\beta}})}{B}\right), \tag{B.10}$$

and the corresponding FOC provides a closed-form solution for the parameter estimates:

$$\nabla_{\boldsymbol{\alpha}}\Lambda(\theta) = 0 = -2\left(\frac{1}{B}\sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{W}_{bt-1}^{T}\boldsymbol{W}_{bt-1}\right)\left(\frac{\boldsymbol{W}_{bt-1}^{T}\boldsymbol{W}_{bt-1}}{B}\right)^{-1}\left(\frac{1}{B}\sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{W}_{\boldsymbol{bt-1}}^{\boldsymbol{T}}\eta_{bt}\right)$$

$$\Leftrightarrow$$

$$\boldsymbol{\alpha} = \left(\left(\frac{\sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{W}_{bt-1}^{T}\boldsymbol{W}_{bt-1}}{B}\right)\left(\frac{\sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{W}_{bt-1}^{T}\boldsymbol{W}_{bt-1}}{B}\right)^{-1}\left(\frac{\sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{W}_{bt-1}^{T}\boldsymbol{W}_{bt-1}}{B}\right)\right)^{-1}$$
$$\times\left(\left(\frac{\sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{W}_{bt-1}^{T}\boldsymbol{W}_{bt-1}}{B}\right)\left(\frac{\sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{W}_{bt-1}^{T}\boldsymbol{W}_{bt-1}}{B}\right)^{-1}\left(\frac{\sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{W}_{bt-1}^{T}\omega_{bt}}{B}\right)\right)$$

$$\Leftrightarrow$$

$$\boldsymbol{\alpha} = \left(\sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{W}_{bt-1}^{T}\boldsymbol{W}_{bt-1}\right)^{-1}\left(\sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{W}_{bt-1}^{T}\omega_{bt}\right)$$

B.2 First-order condition methods

Starting from a gross output production function:

$$y_{bt} = f^{klm}\left(k_{bt}, l_{bt}, m_{bt}\right) + \omega_{bt} + \varepsilon_{bt}, \tag{B.11}$$

the estimator proposed by Gandhi et al. (2020) consists of two stages. In a *first stage*, one relies on the log-linearized material share equation, obtained from the first-order condition for the profit-maximizing decision on material inputs, to identify the elasticity of output with respect to materials and the ex-post Hicks-neutral productivity shock ε_{bt}:

$$log\left(\frac{P_t^M M_{bt}}{P_t^Y Y_{bt}}\right) = log\left(\mathcal{E}\right) + log\left(\frac{\partial f^{klm}\left(k_{bt}, l_{bt}, m_{bt}\right)}{\partial m_{bt}}\right) - \varepsilon_{bt} \tag{B.12}$$

where $\mathcal{E} = E\left[e^{\varepsilon_{bt}}\right]$ and P_t^M, P_t^Y are aggregate material and output prices, respectively. The output from this first stage estimation enables us to define the 'non-flexible' output variation as:

$$\phi_{bt} = y_{bt} - \varepsilon_{bt} - \int \frac{\partial f^{klm}\left(k_{bt}, l_{bt}, m_{bt}\right)}{\partial m_{bt}} dm_{bt} = -f^{kl}\left(k_{bt}, l_{bt}\right) + \omega_{bt}. \tag{B.13}$$

Relying on the productivity evolving according to a first-order Markov process, $\omega_{bt} = g(\omega_{bt-1}) + \eta_{bt}$, this results in the *second stage* estimation equation

$$\phi_{bt} = -f^{kl}\left(k_{bt}, l_{bt}\right) + g\left(\phi_{bt-1} - f^{kl}\left(k_{bt-1}, l_{bt-1}\right)\right) + \eta_{bt}. \tag{B.14}$$

Consistent parameter estimates for the production function can be obtained building on the moment conditions following from the independence between the timing of factor input decisions and the unexpected shocks to productivity:

$$E\left[\eta_{bt}|k_{bt}, l_{bt(-1)}, \phi_{bt-1}\right] = 0. \tag{B.15}$$

We parametrize equation (B.14) with production function coefficients $\boldsymbol{\beta}$ and specify the linear first-order Markov process, $g(\phi_{bt-1}, l_{bt-1}, k_{bt-1}; \boldsymbol{\beta}, \boldsymbol{\alpha}) = \boldsymbol{W}_{bt-1}\boldsymbol{\alpha}$ with $\boldsymbol{W}_{bt-1} = [1, \omega_{bt-1}]$, and $\theta \equiv \{\boldsymbol{\beta}, \boldsymbol{\alpha}\}$ such that

$$\phi_{bt} = -f^{kl}\left(k_{bt}, l_{bt}; \boldsymbol{\beta}\right) + \boldsymbol{W}_{bt-1}\boldsymbol{\alpha} + \eta_{bt}. \tag{B.16}$$

This specification takes a very similar form to the estimation equation for the proxy variable method specified above. The remaining optimization criterion (B.9) and its solution are equivalent to that in the proxy variable methods.

B.3 Mixture (limited information) maximum likelihood

The methodology proposed in this paper builds on existing two-stage estimation methods for the first stage estimation (Ackerberg et al., 2015; Gandhi et al., 2020). These first-stage estimation procedures (see above) are consistent with the proposed generalization of the Markov process of productivity, as they rely on flexible production factors unaffected by different expectations regarding future productivity shocks between groups of firms (Ackerberg, 2021). As discussed in the main text, however, the second-stage specification is dependent on the timing assumption of the labor input decision. We specify the estimator for different timing assumptions below.

B.3.1 Labor as a dynamic input but not predetermined input

If labor is assumed to be a dynamic but not predetermined input, we have to consider the possible correlation between the unexpected shock to productivity and labor choice (Ackerberg et al., 2015). As discussed in the main text, the second-stage estimation equation can then be represented as follows:

$$\mathcal{L}^c\left(\boldsymbol{\Theta}, \boldsymbol{z}\right) = \sum_{b=1}^{B}\sum_{s=1}^{S} z_b^s log\left(Pr(z_b^s|k_{b0}, l_{b0}, \omega_{b0}; \boldsymbol{\gamma}^s)\prod_{t=1}^{T} p(\phi_{bt}, l_{bt}|k_{bt}, \phi_{bt-1}, l_{bt-1}, k_{bt-1}, z_b^s; \boldsymbol{\theta}^s)\right). \tag{B.17}$$

We estimate the parameters of interest based on equation (B.17) 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 by $(\boldsymbol{\Theta})^j \equiv \left\{(\gamma^1)^j, \ldots, (\gamma^S)^j, (\boldsymbol{\theta}^1)^j, \ldots, (\boldsymbol{\theta}^S)^j\right\}$, then the steps of the iterative procedure are as follows:

1. Use the current-iteration starting values for the parameters, $(\boldsymbol{\Theta})^j$,[1] and approximate cluster affiliation with the posterior conditional probability obtained from Bayes' theorem:

$$\hat{z}_b^s = Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; (\boldsymbol{\Theta})^j) = \frac{Pr(z_b^s|k_{b0}, l_{b0}, \omega_{b0}; (\boldsymbol{\gamma}^s)^j)\prod_{t=1}^{T} p(\phi_{bt}, l_{bt}|k_{bt}, \phi_{bt-1}, l_{bt-1}, k_{bt-1}, z_b^s; \boldsymbol{\theta}^s)}{p^o\left(\boldsymbol{\phi}, \boldsymbol{l}; (\boldsymbol{\Theta})^j\right)}. \tag{B.18}$$

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

[1] Starting values for the first iteration are obtained from an OLS production function estimation.

$$
\text{(i)} \quad \begin{aligned} \max_{(\boldsymbol{\theta})^{j+1}} \Lambda(\boldsymbol{\theta}^{j+1}) = \max_{(\boldsymbol{\theta})^{j+1}} & \sum_{b=1}^{B}\sum_{s=1}^{S} Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; (\boldsymbol{\Theta})^j) \\ & \times log\left(\prod_{t=1}^{T} p(\phi_{bt}, l_{bt}|k_{bt}, \phi_{bt-1}, k_{bt-1}, l_{bt-1}, z_b^s; (\boldsymbol{\theta}^s))^{j+1}\right); \end{aligned}
$$

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

The maximum likelihood estimation of the conditional probability of cluster affiliation, step 2.(ii), is implemented using the *multinom* function of the *nnet* R package with maximum likelihood (i.e. when entropy = TRUE) rather than least-squares optimization (i.e. when entropy = FALSE). However, the maximum likelihood estimation of the cluster-probability weighted observed log-likelihood ($\Lambda(\theta^{j+1})$), step 2.(i), is slightly more involved.

As specified in the main text, the observed likelihood attains a bivariate normal specification when conditioning on instrumental variables for endogenous regressors:

$$
p^o(\phi_{bt}, l_{bt}|k_{bt}, \phi_{bt-1}, k_{bt-1}, l_{bt-1}, z_b^s; \underbrace{\boldsymbol{\beta}, \boldsymbol{\alpha}^s, \boldsymbol{\delta}^s, \boldsymbol{\Sigma}^s}_{\equiv \boldsymbol{\theta}^s}) = \frac{e^{-\frac{1}{2}(\boldsymbol{\epsilon}^s)^T(\boldsymbol{\Sigma}^s)^{-1}(\boldsymbol{\epsilon}^s)}}{\sqrt{(2\pi)^2|\boldsymbol{\Sigma}^s|}}, \tag{B.19}
$$

where $\boldsymbol{\epsilon}^s = \begin{bmatrix} \phi_{bt} - f^{kl}(k_{bt}, l_{bt}; \boldsymbol{\beta}) - g(\phi_{bt-1}, k_{bt-1}, l_{bt-1}; \boldsymbol{\beta}, \boldsymbol{\alpha^s}) \\ l_{bt} - \delta_0^s - \delta_1 k_{bt} - \delta_2^s \phi_{bt-1} - \delta_3^s k_{bt-1} - \delta_4^s l_{bt-1} \end{bmatrix}$ and $\boldsymbol{\Sigma}^s = \begin{bmatrix} (\sigma_\eta^s)^2 & \sigma_{\eta,\zeta}^s \\ \sigma_{\eta,\zeta}^s & (\sigma_\zeta^s)^2 \end{bmatrix}$.

To simplify the estimation procedure, we rely on the observation that equation (B.19) can be factorized into a density of the endogenous variables conditional on the instrumental variables, $p^o(\phi_{bt}, l_{bt}) = p^o(\phi_{bt}|l_{bt})p^o(l_{bt})$, such that

$$
p^o(\phi_{bt}|k_{bt}, l_{bt}, \phi_{bt-1}, k_{bt-1}, l_{bt-1}, z_b^s; \boldsymbol{\beta}, \boldsymbol{\alpha}^s, \sigma_\eta^s, \sigma_{\eta,\zeta}^s) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\left[\left(\sigma_\eta^s\right)^2 - \frac{(\sigma_{\eta,\zeta}^s)^2}{\left(\hat{\sigma}_\zeta^s\right)^2}\right]}} e^{-\frac{1}{2}\frac{\left(\eta_{bt}^s - \frac{\sigma_{\eta,\zeta}^s}{\left(\hat{\sigma}_\zeta^s\right)^2}\zeta_{bt}^s\right)^2}{(\sigma_\eta^s)^2 - \frac{\sigma_{\eta,\zeta}^s}{\left(\hat{\sigma}_\zeta^s\right)^2}}}, \tag{B.20}
$$

and

$$
p^o(l_{bt}|k_{bt}, \phi_{bt-1}, k_{bt-1}, l_{bt-1}, z_b^s; \boldsymbol{\delta}^s, \sigma_\zeta^s) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\left(\sigma_\zeta^s\right)^2}} e^{-\frac{1}{2}\left(\frac{\zeta_{bt}^s}{\sigma_\zeta^s}\right)^2}. \tag{B.21}
$$

We then rely on a two-step procedure to obtain the MLE estimates (see, for instance, Kutlu, 2010). In the first step, we gather the instrumental variables in the column vector $\boldsymbol{Z}$ and obtain the parameters of the reduced-form equation from the first-order condition (FOC):

1. $\nabla_{\boldsymbol{\delta}^s}\Lambda(\boldsymbol{\theta}) = 0 = -\frac{1}{\left(\sigma_\zeta^s\right)^2}\sum_{b=1}^{B} Pr(z_b^s|\boldsymbol{k}_b,\boldsymbol{l}_b,\boldsymbol{\phi}_b;\boldsymbol{\Theta})\sum_{t=1}^{T}\boldsymbol{Z}_{bt}^T\left(l_{bt}-\boldsymbol{Z}_{bt}\boldsymbol{\delta}^s\right)$

$$\Leftrightarrow$$

$$0 = \sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{Z}_{bt}^T Pr(z_b^s|\boldsymbol{k}_b,\boldsymbol{l}_b,\boldsymbol{\phi}_b;\boldsymbol{\Theta})\left(l_{bt}-\boldsymbol{Z}_{bt}\boldsymbol{\delta}^s\right)$$

$$\boldsymbol{\delta}^s = \left(\sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{Z}_{bt}^T Pr(z_b^s|\boldsymbol{k}_b,\boldsymbol{l}_b,\boldsymbol{\phi}_b;\boldsymbol{\Theta})\boldsymbol{Z}_{bt}\right)^{-1}\sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{Z}_{bt}^T Pr(z_b^s|\boldsymbol{k}_b,\boldsymbol{l}_b,\boldsymbol{\phi}_b;\boldsymbol{\Theta}) l_{bt};$$

2. $\left(\hat{\sigma}_\zeta^s\right)^2 = \frac{\sum_{b=1}^{B}\sum_{t=1}^{T} Pr(z_b^s|\boldsymbol{k}_b,\boldsymbol{l}_b,\boldsymbol{\phi}_b;\boldsymbol{\Theta})\left(\zeta_{bt}^s\right)^2}{\sum_{b=1}^{B}\sum_{t=1}^{T} Pr(z_b^s|\boldsymbol{k}_b,\boldsymbol{l}_b,\boldsymbol{\phi}_b;\boldsymbol{\Theta})};$

In the second step, we take the parameters obtained in the first step as given and estimate the remaining parameters. We specify the linear first-order Markov process, $g(\phi_{bt-1},k_{bt-1},l_{bt-1};\boldsymbol{\beta},\boldsymbol{\alpha^s}) = \boldsymbol{W}_{bt-1}\boldsymbol{\alpha}^s$ with $\boldsymbol{W}_{bt-1}=[1,\omega_{bt-1}]$. The log-likelihood is linear in the parameters $\boldsymbol{\alpha}^s$ and non-linear in the parameters $\boldsymbol{\beta}$, leading to the following optimization conditions:

3. $\nabla_{\boldsymbol{\alpha}^s}\Lambda(\boldsymbol{\theta}) = 0 = -\frac{1}{\left(\sigma_\eta^s\right)^2-\frac{(\sigma_{\eta,\zeta}^s)^2}{\left(\sigma_\zeta^s\right)^2}}\sum_{b=1}^{B} Pr(z_b^s|\boldsymbol{k}_b,\boldsymbol{l}_b,\boldsymbol{\phi}_b;\boldsymbol{\Theta})\sum_{t=1}^{T}\left(\nabla_{\boldsymbol{\alpha}^s}\left(\eta_{bt}^s-\frac{\sigma_{\eta,\zeta}^s}{\left(\sigma_\zeta^s\right)^2}\zeta_{bt}^s\right)\right)^T\left(\eta_{bt}^s-\frac{\sigma_{\eta,\zeta}^s}{\left(\sigma_\zeta^s\right)^2}\zeta_{bt}^s\right)$

$$0 = \sum_{b=1}^{B}\sum_{t=1}^{T}\left(\nabla_{\boldsymbol{\alpha}^s}\eta_{bt}^s\right)^T Pr(z_b^s|\boldsymbol{k}_b,\boldsymbol{l}_b,\boldsymbol{\phi}_b;\boldsymbol{\Theta})\left(\eta_{bt}^s-\frac{\sigma_{\eta,\zeta}^s}{\left(\sigma_\zeta^s\right)^2}\zeta_{bt}^s\right)$$

$$0 = \sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{W}_{bt-1}^T Pr(z_b^s|\boldsymbol{k}_b,\boldsymbol{l}_b,\boldsymbol{\phi}_b;\boldsymbol{\Theta})\left(\omega_{bt}-\boldsymbol{W}_{bt-1}\boldsymbol{\alpha}^s-\frac{\sigma_{\eta,\zeta}^s}{\left(\sigma_\zeta^s\right)^2}\zeta_{bt}^s\right)$$

$$0 = \sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{W}_{bt-1}^T Pr(z_b^s|\boldsymbol{k}_b,\boldsymbol{l}_b,\boldsymbol{\phi}_b;\boldsymbol{\Theta})\left(\omega_{bt}-\frac{\sigma_{\omega,\zeta}^s}{\left(\sigma_\zeta^s\right)^2}\zeta_{bt}^s-\left[\boldsymbol{W}_{bt-1}-\frac{\sigma_{\boldsymbol{W}_{bt-1},\zeta}}{\left(\sigma_\zeta^s\right)^2}\zeta_{bt}^s\right]\boldsymbol{\alpha}^s\right)$$

$$\boldsymbol{\alpha}^s = \left(\sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{W}_{bt-1}^T Pr(z_b^s|\boldsymbol{k}_b,\boldsymbol{l}_b,\boldsymbol{\phi}_b;\boldsymbol{\Theta})\left(\boldsymbol{W}_{bt-1}-\frac{\sigma_{\boldsymbol{W}_{bt-1},\zeta}^s}{\left(\sigma_\zeta^s\right)^2}\zeta_{bt}^s\right)\right)^{-1}$$

$$\times\sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{W}_{bt-1}^T Pr(z_b^s|\boldsymbol{k}_b,\boldsymbol{l}_b,\boldsymbol{\phi}_b;\boldsymbol{\Theta})\left(\omega_{bt}-\frac{\sigma_{\omega,\zeta}^s}{\left(\sigma_\zeta^s\right)^2}\zeta_{bt}^s\right)$$

4. $\nabla_{\boldsymbol{\beta}}\Lambda(\boldsymbol{\theta}) = 0 = \sum_{s=1}^{S}-\frac{1}{\left(\sigma_\eta^s\right)^2-\frac{(\sigma_{\eta,\zeta}^s)^2}{\left(\sigma_\zeta^s\right)^2}}\sum_{b=1}^{B} Pr(z_b^s|\boldsymbol{k}_b,\boldsymbol{l}_b,\boldsymbol{\phi}_b;\boldsymbol{\Theta})\sum_{t=1}^{T}\left(\nabla_{\boldsymbol{\beta}}\eta_{bt}^s\right)^T\left(\eta_{bt}^s-\frac{\sigma_{\eta,\zeta}^s}{\left(\sigma_\zeta^s\right)^2}\zeta_{bt}^s\right)$

$$\Leftrightarrow$$

$$0 = \sum_{s=1}^{S}\sum_{b=1}^{B}\sum_{t=1}^{T}\frac{1}{\left(\sigma_\eta^s\right)^2-\frac{\sigma_{\eta,\zeta}^s}{\left(\sigma_\zeta^s\right)^2}} Pr(z_b^s|\boldsymbol{k}_b,\boldsymbol{l}_b,\boldsymbol{\phi}_b;\boldsymbol{\Theta})\left(\nabla_{\boldsymbol{\beta}}(\eta_{bt}^s)\right)^T\left(\eta_{bt}^s-\frac{\sigma_{\eta,\zeta}^s}{\left(\sigma_\zeta^s\right)^2}\zeta_{bt}^s\right),$$

where $\nabla_{\boldsymbol{\beta}}(\eta_{bt}^s) = -\nabla_{\boldsymbol{\beta}}f^{kl}(k_{bt},l_{bt};\boldsymbol{\beta})+\alpha_2^s\nabla_{\boldsymbol{\beta}}f^{kl}(k_{bt-1},l_{bt-1};\boldsymbol{\beta})+\boldsymbol{W}_{bt-1}\nabla_{\boldsymbol{\beta}}\alpha_2^s$.

5. $\left(\hat{\sigma}_\eta^s\right)^2 = \frac{\sum_{b=1}^{B}\sum_{t=1}^{T} Pr(z_b^s|\boldsymbol{k}_b,\boldsymbol{l}_b,\boldsymbol{\phi}_b;\boldsymbol{\Theta})\left(\eta_{bt}^s\right)^2}{\sum_{b=1}^{B}\sum_{t=1}^{T} Pr(z_b^s|\boldsymbol{k}_b,\boldsymbol{l}_b,\boldsymbol{\phi}_b;\boldsymbol{\Theta})};$

6. $\hat{\sigma}^s_{\eta,\zeta} = \dfrac{\sum_{b=1}^B \sum_{t=1}^T Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \boldsymbol{\Theta})\eta^s_{bt}\zeta^s_{bt}}{\sum_{b=1}^B \sum_{t=1}^T Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \boldsymbol{\Theta})}.$

Notice that this two-step procedure is essentially a control function approach (Amsler et al., 2016) that allows us to obtain all cluster-specific parameters based on a closed-form solution despite the non-linearity of the overall optimization problem. Moreover, the dimension of the non-linear optimization problem becomes independent of the number of clusters and significantly reduces the additional computational time needed when increasing the number of clusters.

B.3.2 Labor as a predetermined input

If labor is assumed to be predetermined, there are no endogeneity concerns in the second stage of the estimation, and the observed likelihood can be specified as a univariate normal distribution:

$$p^o(\phi_{bt}|k_{bt}, l_{bt}, \phi_{bt-1}, l_{bt-1}, k_{bt-1}, z_b^s; \boldsymbol{\beta}, \boldsymbol{\alpha}^s, \sigma^s_\eta) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi\left(\sigma^s_\eta\right)^2}} e^{-\frac{1}{2}\left(\frac{\eta^s_{bt}}{\sigma^s_\eta}\right)^2}. \tag{B.22}$$

The FOC are then:

1. $$\nabla_{\boldsymbol{\alpha}^s}\Lambda(\boldsymbol{\theta}) = 0 = -\frac{1}{\left(\sigma^s_\eta\right)^2}\sum_{b=1}^B Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \boldsymbol{\Theta}) \sum_{t=1}^T \boldsymbol{W}^T_{bt-1}\left(\omega_{bt} - \boldsymbol{W}_{bt-1}\boldsymbol{\alpha}^s\right)$$
$$\Leftrightarrow$$
$$0 = \sum_{b=1}^B \sum_{t=1}^T \boldsymbol{W}^T_{bt-1} Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \boldsymbol{\Theta})\left(\omega_{bt} - \boldsymbol{W}_{bt-1}\boldsymbol{\alpha}^s\right)$$
$$\boldsymbol{\alpha}^s = \left(\sum_{b=1}^B \sum_{t=1}^T \boldsymbol{W}^T_{bt-1} Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \boldsymbol{\Theta})\boldsymbol{W}_{bt-1}\right)^{-1} \sum_{b=1}^B \sum_{t=1}^T \boldsymbol{W}^T_{bt-1} Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \boldsymbol{\Theta})\omega_{bt};$$
2. $$\nabla_{\boldsymbol{\beta}}\Lambda(\boldsymbol{\theta}) = 0 = \sum_{s=1}^S -\frac{1}{\left(\sigma^s_\eta\right)^2}\sum_{b=1}^B Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \boldsymbol{\Theta}) \sum_{t=1}^T \nabla_{\boldsymbol{\beta}}(\eta^s_{bt})\left(\omega_{bt} - \boldsymbol{W}_{bt-1}\boldsymbol{\alpha}^s\right)$$
$$\Leftrightarrow$$
$$0 = \sum_{s=1}^S \sum_{b=1}^B \sum_{t=1}^T \frac{1}{\left(\sigma^s_\eta\right)^2}\nabla_{\boldsymbol{\beta}}\left(\eta^s_{bt}\right) Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \boldsymbol{\Theta})\left(\omega_{bt} - \boldsymbol{W}_{bt-1}\boldsymbol{\alpha}^s\right),$$
where $\nabla_{\boldsymbol{\beta}}(\eta^s_{bt}) = -\nabla_{\boldsymbol{\beta}} f^{kl}\left(k_{bt}, l_{bt}; \boldsymbol{\beta}\right) + \alpha^s_2 \nabla_{\boldsymbol{\beta}} f^{kl}\left(k_{bt-1}, l_{bt-1}; \boldsymbol{\beta}\right) - \boldsymbol{W}_{bt-1}\nabla_{\boldsymbol{\beta}}\alpha^s_2;$
3. $$\frac{\partial\Lambda(\boldsymbol{\theta})}{\partial\left(\sigma^s_\eta\right)^2} = 0 = \sum_{b=1}^B Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \boldsymbol{\Theta})\left(-\frac{T}{2\left(\sigma^s_\eta\right)^2} + \frac{1}{2(\sigma^s_\eta)^4}\sum_{t=1}^T (\eta^s_{bt})^2\right)$$
$$\Leftrightarrow$$
$$\left(\hat{\sigma}^s_\eta\right)^2 = \frac{\sum_{b=1}^B \sum_{t=1}^T Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \boldsymbol{\Theta})\left(\eta^s_{bt}\right)^2}{\sum_{b=1}^B \sum_{t=1}^T Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \boldsymbol{\Theta})}.$$

Appendix C Monte Carlo

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

Production function and productivity shocks.— We simulate a panel dataset of 1,000 firms over 10 years assuming a Leontief production function:

$$Y_{bt} = min\left\{K_{bt}^{\beta_k} L_{bt}^{\beta_l} e^{\omega_{bt}}, \beta_m M_{bt}\right\} e^{\varepsilon_{bt}} \tag{C.1}$$

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

$$\omega_{bt} = \sum_{s=1}^{2} [\alpha_0^s + \alpha_1^s \omega_{bt-1} + \eta_{bt}^s]^{z_b^s}, \tag{C.2}$$

with 800 observations assigned to cluster one ($s = 1$), $Pr(z_b^1) = 0.8$, and 200 observations to cluster two ($s = 2$), $Pr(z_b^2) = 0.2$. We assume that the cluster-specific unexpected shocks to productivity follow a normal distribution, $\eta_{bt}^s \sim \mathcal{N}\left(0, \sigma_\eta^s\right)$.

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, though labor is predetermined (see Section 3.1). L_{bt} is chosen prior to period t without full knowledge of ω_{bt}. Strictly speaking, labor is chosen at time $t - i$, with $i = 0.5$. We can think of decomposing the finite mixture AR(1)-process (C.2) into two sub-processes. First, ω_{bt-1} evolves to $\omega_{b,t-i}$, at which point in time the firm chooses its labor input (as a function of $\omega_{b,t-i}$). After L_{bt} is chosen, $\omega_{b,t-i}$ evolves to ω_{bt}. Additionally, there are firm-specific (unobserved to the econometrician) wage shocks.

The evolution of ω between sub-periods is specified as follows:

$$\begin{aligned}
\omega_{b,t-i} &= \sum_{s=1}^{2} \left[\alpha_0^s + (\alpha_1^s)^{1-i} \omega_{b,t-1} + \eta_{bt}^{c,A}\right]^{z_b^s}; \\
\omega_{bt} &= \sum_{c=1}^{2} \left[\left(1 - (\alpha_1^s)^i\right) \alpha_0^s + (\alpha_1^s)^i \omega_{b,t-i} + \eta_{bt}^{c,B}\right]^{z_b^s}.
\end{aligned} \tag{C.3}$$

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

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

$$\ln\left(W_{bt}\right) = 0.3\ln\left(W_{bt-1}\right) + \eta_{bt}^W, \tag{C.4}$$

where the variance of the normally distributed innovation $\eta_{bt}^W\left(\sigma^2_{\eta^W}\right)$ and the initial value $\ln\left(W_{b0}\right)$ are set in such a way that the standard deviation of $\ln\left(W_{bt}\right)$ 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 $\ln\left(L_{bt}\right)$ by about 10% (Ackerberg et al., 2015).

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

$$L_{bt} = \beta_l^{1/(1-\beta_l)} W_{bt}^{-1/(1-\beta_l)} K_{bi}^{\beta_k/(1-\beta_l)} e^{(1/(1-\beta_l))\left(\left(1-(\alpha_1^s)^i\right)\alpha_0^s + \left(\alpha_1^s\right)^i \omega_{bt-1} + (1/2)\sigma^2_{\eta^{s,B}}\right)},$$

for which we rely on the analytical result for the first moment of a log-normally distributed variable, $E_{t-i}\left[e^{\omega_{bt}}\right] = e^{\left(1-(\alpha_1^s)^i\right)\alpha_0^s + \left(\alpha_1^s\right)^i \omega_{b,t-i} + \frac{1}{2}(\eta_{bt}^{s,B})^2}$.

Investment choice and steady state.— In contrast to the flexible labor and material inputs, capital is assumed to be dynamic and accumulated through investment according to $K_{bt} = (1-\delta)K_{bt-1} + I_{bt-1}$, where $(1-\delta) = 0.8$. Investment is subject to convex adjustment costs given by $c_b\left(I_{bt}\right) = \frac{\phi_b}{2} I_{bt}^2$, where $1/\phi_b$ is distributed lognormally across firms (but constant over time) with a standard deviation of 0.6.

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

$$\begin{aligned} I_{bt} = &\frac{\beta}{\phi_b}\sum_{\tau=0}^{\infty}(\beta(1-\delta))^{\tau}\left(\frac{\beta_k}{1-\beta_l}\right) \times \left[\beta_l^{\beta_l/(1-\beta_l)} - \beta_l^{1/(1-\beta_l)}\right] \\ &\times \exp\left\{\left[\left(\frac{1}{1-\beta_l}\right)\alpha_0^c + \left(\frac{1}{1-\beta_l}\right)(\alpha_1^c)^{\tau+1}\omega_{bt} + \frac{-\beta_l}{1-\beta_l}\rho_W^{\tau+1}\ln\left(W_{bt}\right)\right.\right. \\ &+ \frac{1}{2}\left(\frac{-\beta_l}{1-\beta_l}\right)^2 \sigma^2_{\eta^W}\sum_{s=0}^{\tau}\rho_W^{2(\tau-s)} \\ &\left.\left. + \frac{1}{2}\left(\frac{1}{1-\beta_l}\right)^2 (\alpha_1^c)^{2b}\left((\alpha_1^c)^{2\tau}\sigma^2_{\eta^{c,A}} + \sum_{s=1}^{\tau}\rho^{2(\tau-s)}\sigma^2_{\eta}\right) + \left(\frac{1}{1-\beta_l}\right)\left(\frac{1}{2}\sigma^2_{\eta^{c,B}}\right)\right]\right\} \end{aligned} \tag{C.5}$$

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

Appendix D Robustness

D.1 Estimation methodology and cluster selection

The main text reports the estimation results for a value-added Translog production function of NACE Rev.2 industry 22. We demonstrate that the reported results are robust to other estimation methodologies and manufacturing industries. We present the goodness-of-fit indicators (Figure D.1), output elasticities and RTS (Appendix Figures D.2 and D.3), and measures of heterogeneity in the productivity distribution (Appendix Figures D.4 and D.5) for alternative estimation methodologies and all five industries in the data.

The proposed method delivers reasonable estimates in all cases. Moreover, the stability of the output elasticities does not appear to rely on the estimation methodology or any selected industry. The Cobb-Douglas specifications are more volatile than the Translog specifications, but this seems to originate from the model misspecification or local maxima rather than from the underlying heterogeneity in the data. Only the value-added Translog specifications for the entire manufacturing sector and industry 28 demonstrate some signs of an omitted variable bias. However, the estimation results of the respective gross-output Translog specifications do not confirm this observation. Moreover, the stability of the proposed estimator to the addition of supplementary firm-level characteristics is also robust across industries.

In Table D.1 we present the log-likelihood, BIC, and ICLbic for different specifications of the cluster affiliation probabilities for all industries. We observe that, regardless of the industry, the base specification is preferred over a specification with additional firm-level characteristics and that these additional firm-level characteristics are insufficient to account for the uncovered unobserved heterogeneity in productivity.

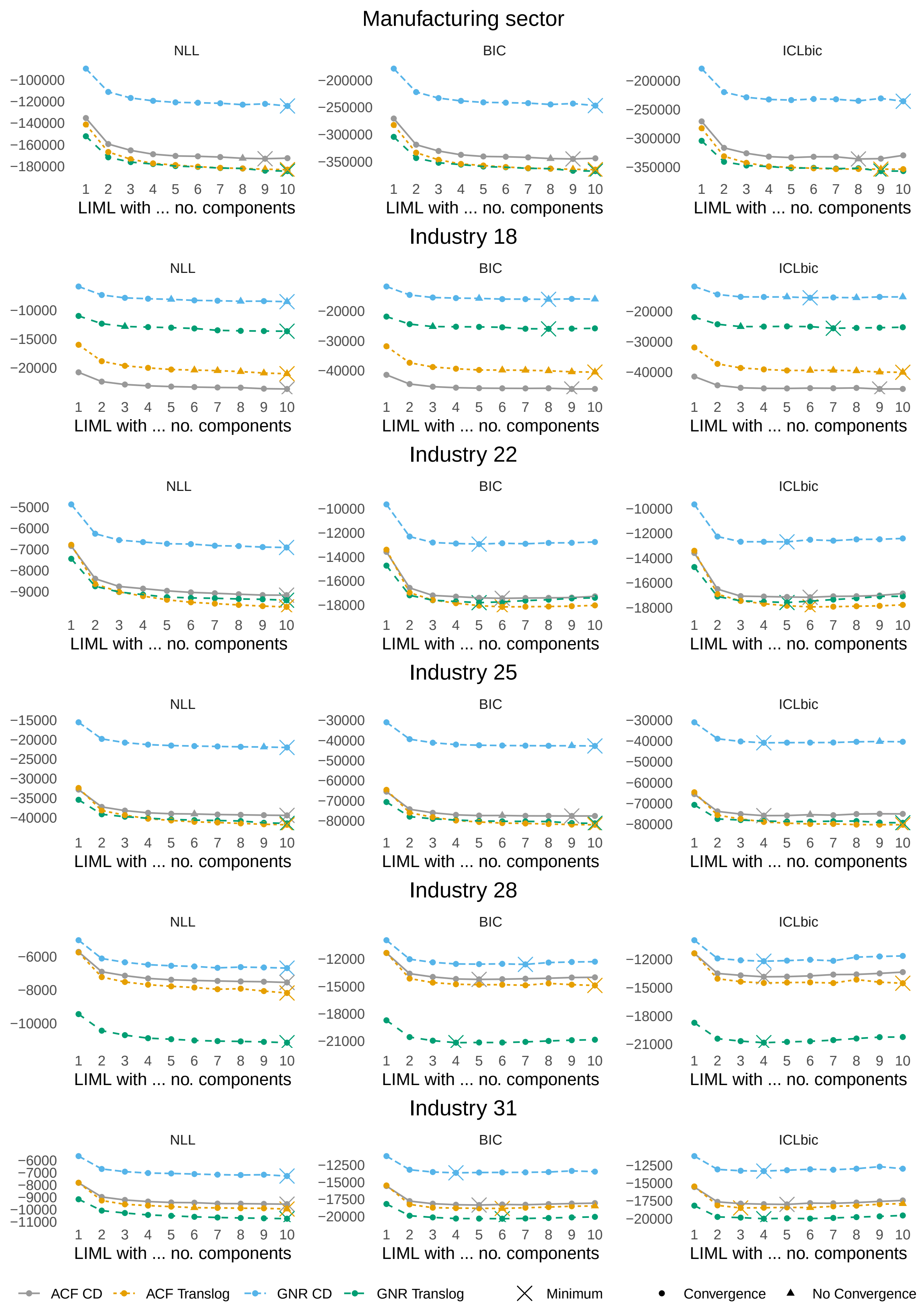

Figure D.1: Change in goodness-of-fit indicators of Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with endogenous labor in function of the number of clusters for the entire manufacturing sector and industries 18, 22, 25, 28, and 31 of the Belgian economy.

Note: NLL stands for negative log-likelihood, BIC for the Bayesian information criterion, and ICLbic for the integrated complete-data likelihood Bayesian information criterion. The times symbol indicates the optimal number of clusters defined by the minimum of the respective goodness-of-fit indicator. "No convergence" indicates non-convergence of the maximum likelihood estimation algorithm.

Figure D.2: Change in output elasticities, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with endogenous labor, in function of the number of clusters for the entire manufacturing sector and industries 18 and 22 of the Belgian economy.

Note: GMM and LIML refer to the generalized method of moments and limited information maximum likelihood as estimation procedures. The times symbol indicates the optimal number of clusters defined by the integrated complete-data likelihood Bayesian information criterion. “No convergence” indicates non-convergence of the maximum likelihood estimation algorithm.

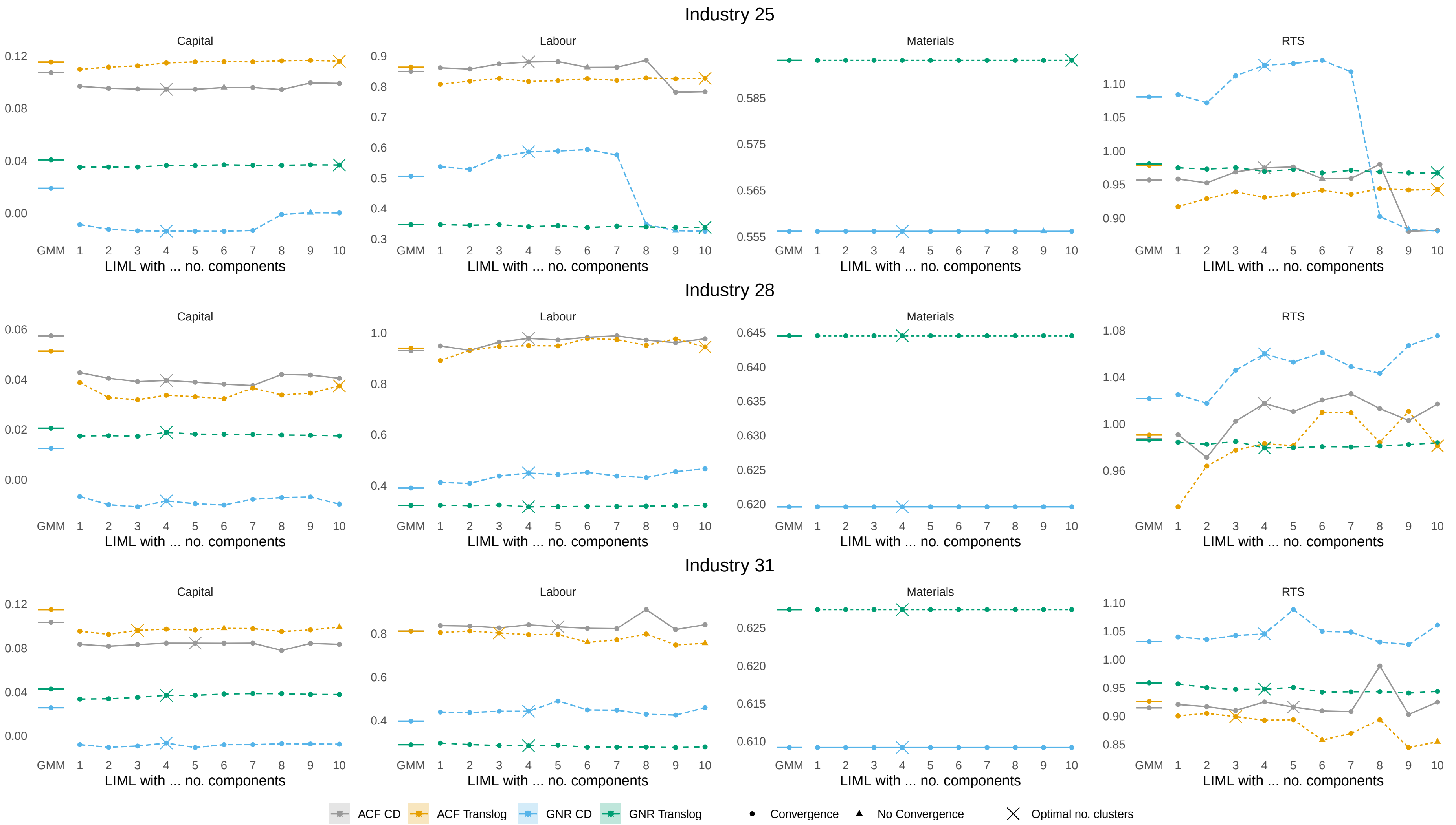

Figure D.3: Change in output elasticities, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with endogenous labor, in function of the number of clusters for industries 25, 28, and 31 of the Belgian economy.

Note: GMM and LIML refer to the generalized method of moments and limited information maximum likelihood as estimation procedure. The times symbol indicates the optimal number of clusters defined by the integrated complete-data likelihood Bayesian information criterion. "No convergence" indicates non-convergence of the maximum likelihood estimation algorithm.

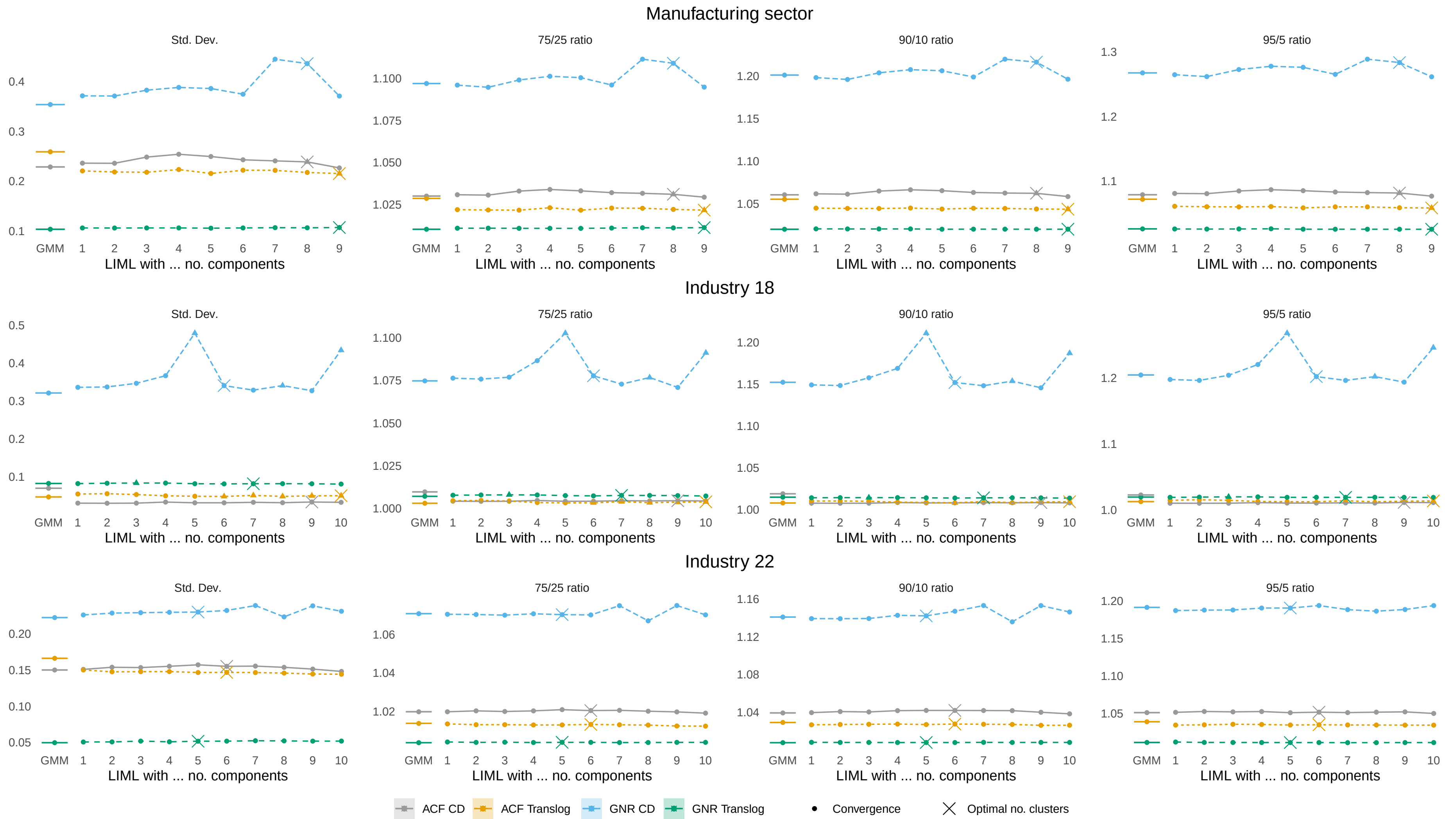

Figure D.4: Change in the standard deviation, 75/25-, 90/10-, and 95/5-ratio, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with endogenous labor, in function of the number of clusters for the entire manufacturing sector and industries 18 and 22 of the Belgian economy.

Note: GMM and LIML refer to the generalized method of moments and limited information maximum likelihood estimation procedures. The times symbol indicates the optimal number of clusters defined by the integrated complete-data likelihood Bayesian information criterion. “No convergence” indicates non-convergence of the maximum likelihood estimation algorithm.

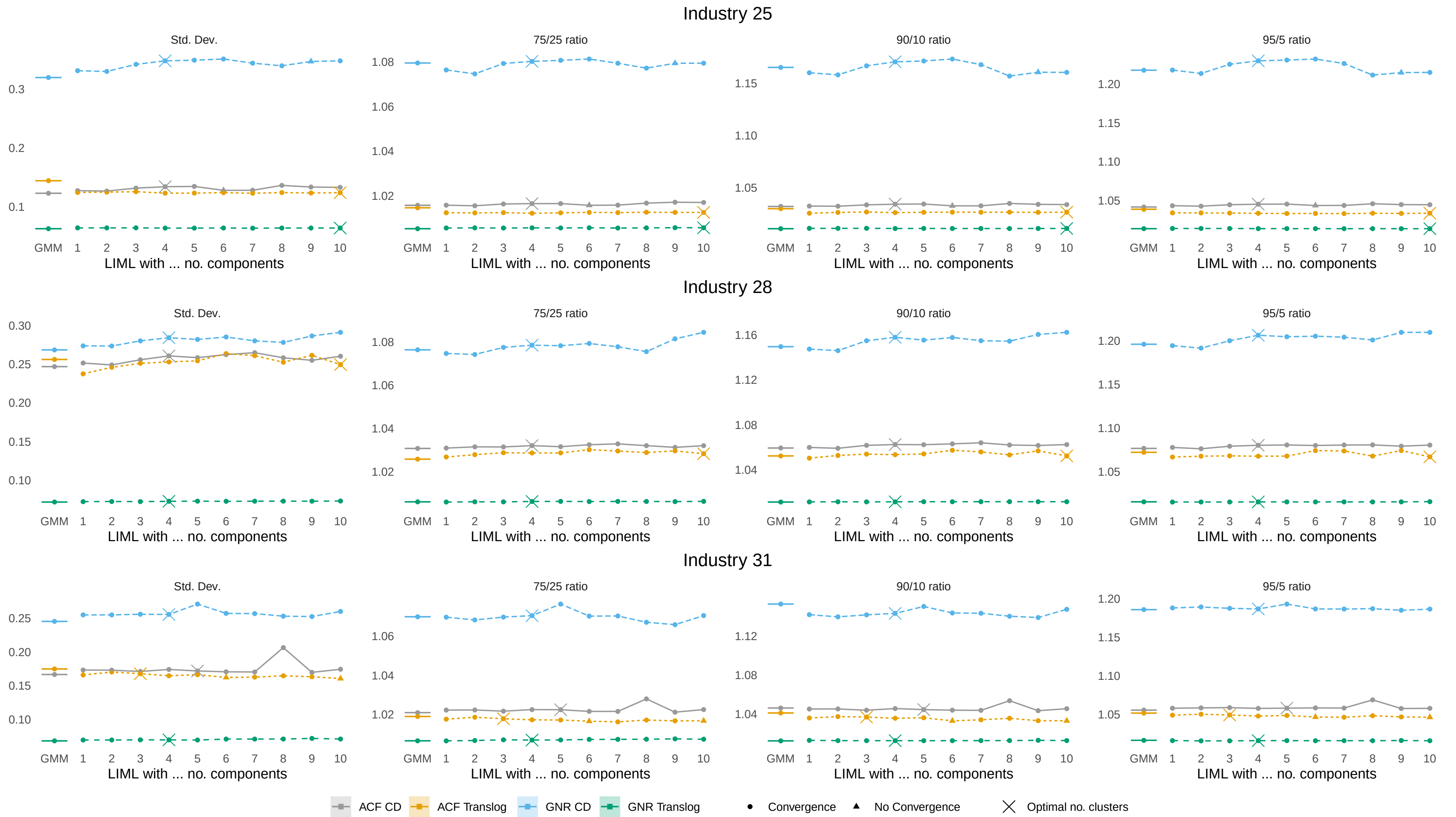

Figure D.5: Change in the standard deviation, 75/25-, 90/10-, and 95/5-ratio, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with endogenous labor, in function of the number of clusters for industries 25, 28, and 31 of the Belgian economy.

Note: GMM and LIML refer to the generalized method of moments and limited information maximum likelihood estimation procedures. The times symbol indicates the optimal number of clusters defined by the integrated complete-data likelihood Bayesian information criterion. “No convergence” indicates non-convergence of the maximum likelihood estimation algorithm.

Table D.1: Goodness-of-fit indicators for estimation with varying concomitant specifications

Specification	Log-likelihood	BIC	ICLbic
Industry 18			
Base specification	21,053.99	-40,602.13	-40,115.06
Additional concomitants	21,101.29	-40,383.38	-39,922.39
Without initial capital and labor	20,462.87	-39,654.90	-38,904.40
Industry 22			
Base specification	9,504.57	-18,150.39	-17,956.60
Additional concomitants	9,515.91	-18,009.49	-17,819.55
Without initial capital and labor	9,291.26	-17,846.43	-17,591.01
Industry 25			
Base specification	41,903.67	-82,116.84	-80,355.24
Additional concomitants	41,991.85	-81,941.42	-80,126.77
Without initial capital and labor	40,850.58	-80,274.49	-77,877.64
Industry 28			
Base specification	8,187.08	-14,916.58	-14,536.71
Additional concomitants	8,223.96	-14,687.03	-14,336.19
Without initial capital and labor	7,797.44	-14,364.79	-13,656.31
Industry 31			
Base specification	9,592.36	-18,729.29	-18,516.11
Additional concomitants	9,601.39	-18,679.87	-18,473.56
Without initial capital and labor	9,464.77	-18,524.71	-18,287.81

Notes: a. The base specification refers to eq. (14), the augmented specification refers to eq. (29), and the specification without initial capital and labor refers to eq. (29) without initial capital and labor.
b. BIC stands for the Bayesian information criterion and ICLbic for the integrated complete-data likelihood Bayesian information criterion.

D.2 Chilean manufacturing sector

To evaluate the generalizability of the proposed productivity estimation methodology and the robustness of the reported results for the Belgian manufacturing sector, we apply our estimation procedure to data on the Chilean manufacturing sector between 1979 and 1996, also used in (Gandhi et al., 2020) and sourced from (Gandhi et al., 2020a). In line with the main results, the goodness of fit statistics displayed in Figure D.8 provide evidence in favor of heterogeneity in productivity. The production function estimates shown in Figures D.9 and D.10 are close to those obtained with current state-of-the-art estimation methodologies. In addition, the shape of the productivity distribution is not significantly affected when increasingly allowing for heterogeneity in productivity, as can be observed from the overall densities in Figures D.6 and D.7, and the summary statistics of the productivity distribution displayed in Figures D.11 and D.12.

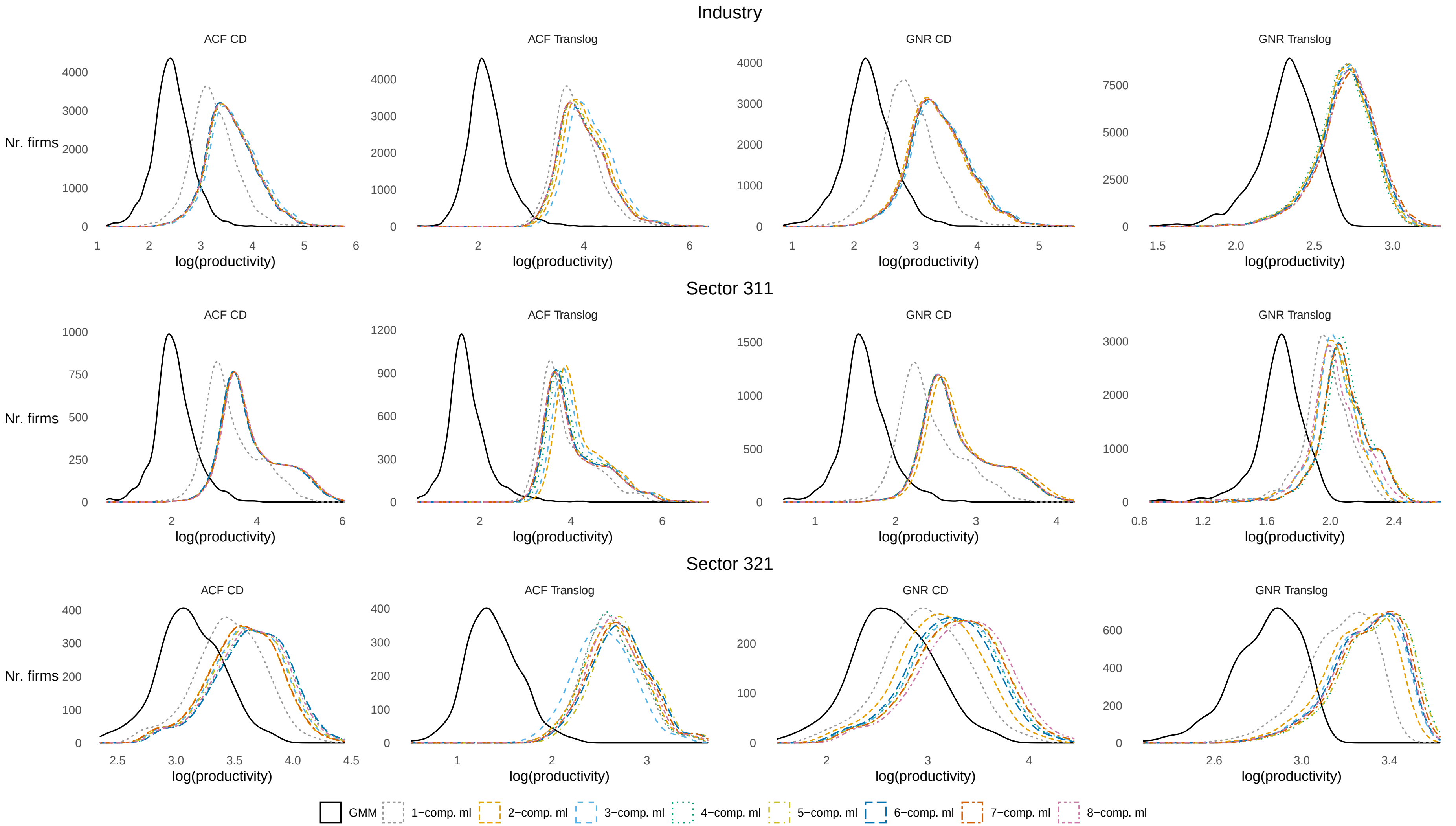

Figure D.6: Density of 1- to 10-cluster productivity in 2013 obtained from Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with endogenous labor for the entire manufacturing sector and industries 311, and 321 of the Chilean economy.

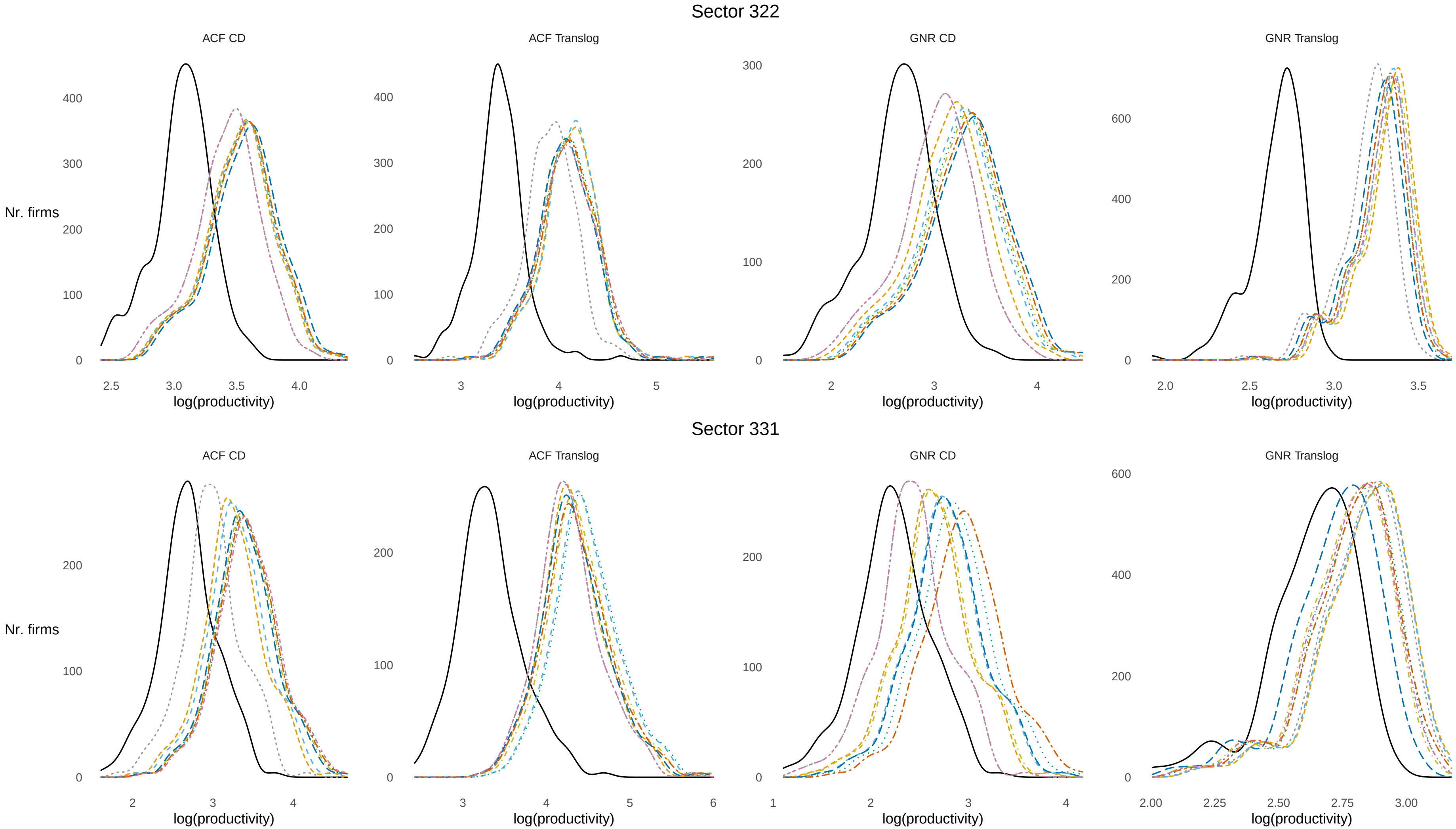

Figure D.7: Density of 1- to 10-cluster productivity in 2013 obtained from Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with endogenous labor for industries 322, 331, and 381 of the Chilean economy.

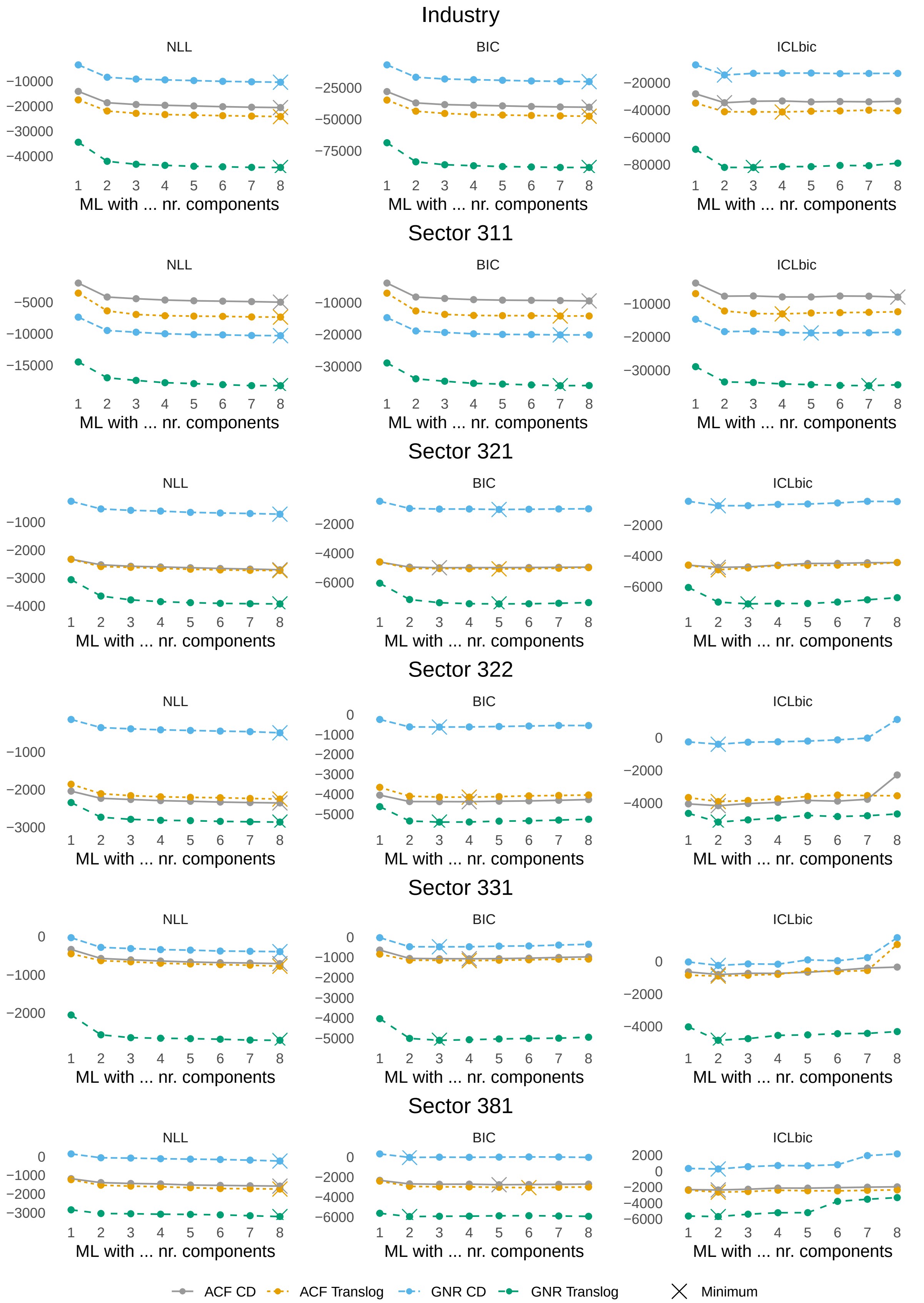

Figure D.8: Change of goodness-of-fit indicators of Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with exogenous labor, in function of the number of clusters for the complete Industry and industries 311, 321, 322, 331, and 381 of the Chilean economy.

Note: NLL stands for negative log-likelihood, BIC for the Bayesian information criterion, and ICLbic for the integrated complete-data likelihood Bayesian information criterion. The times symbol indicates the optimal number of clusters defined by the minimum of the respective goodness-of-fit indicator.

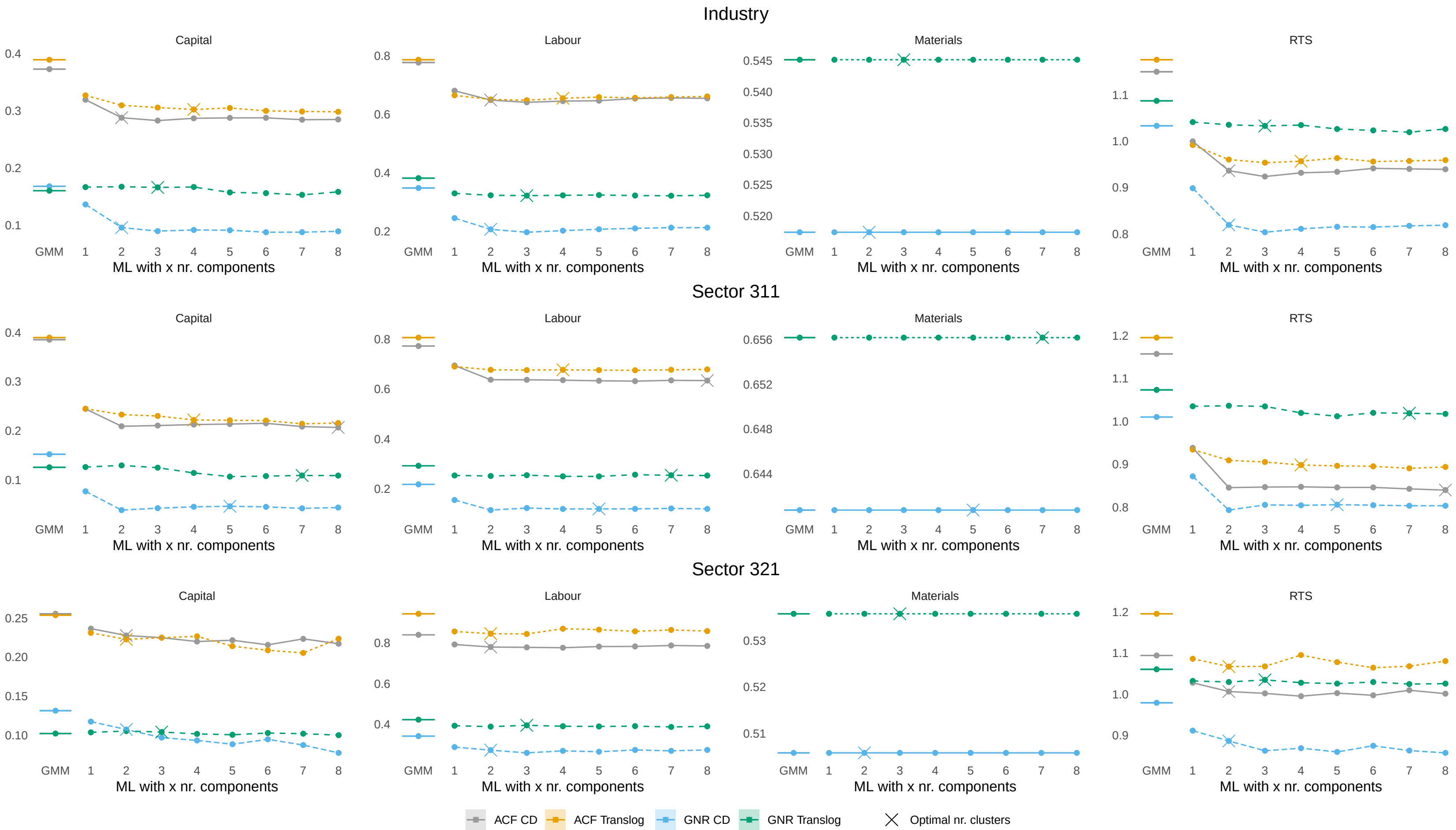

Figure D.9: Evolution of output elasticities, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with exogenous labor, in function of the number of clusters for the complete Industry and industries 311, 321 of the Chilean economy.

Note: GMM and ML refer to the generalized method of moments and maximum likelihood estimation procedures. The times symbol indicates the optimal number of clusters defined by the integrated complete-data likelihood Bayesian information criterion.

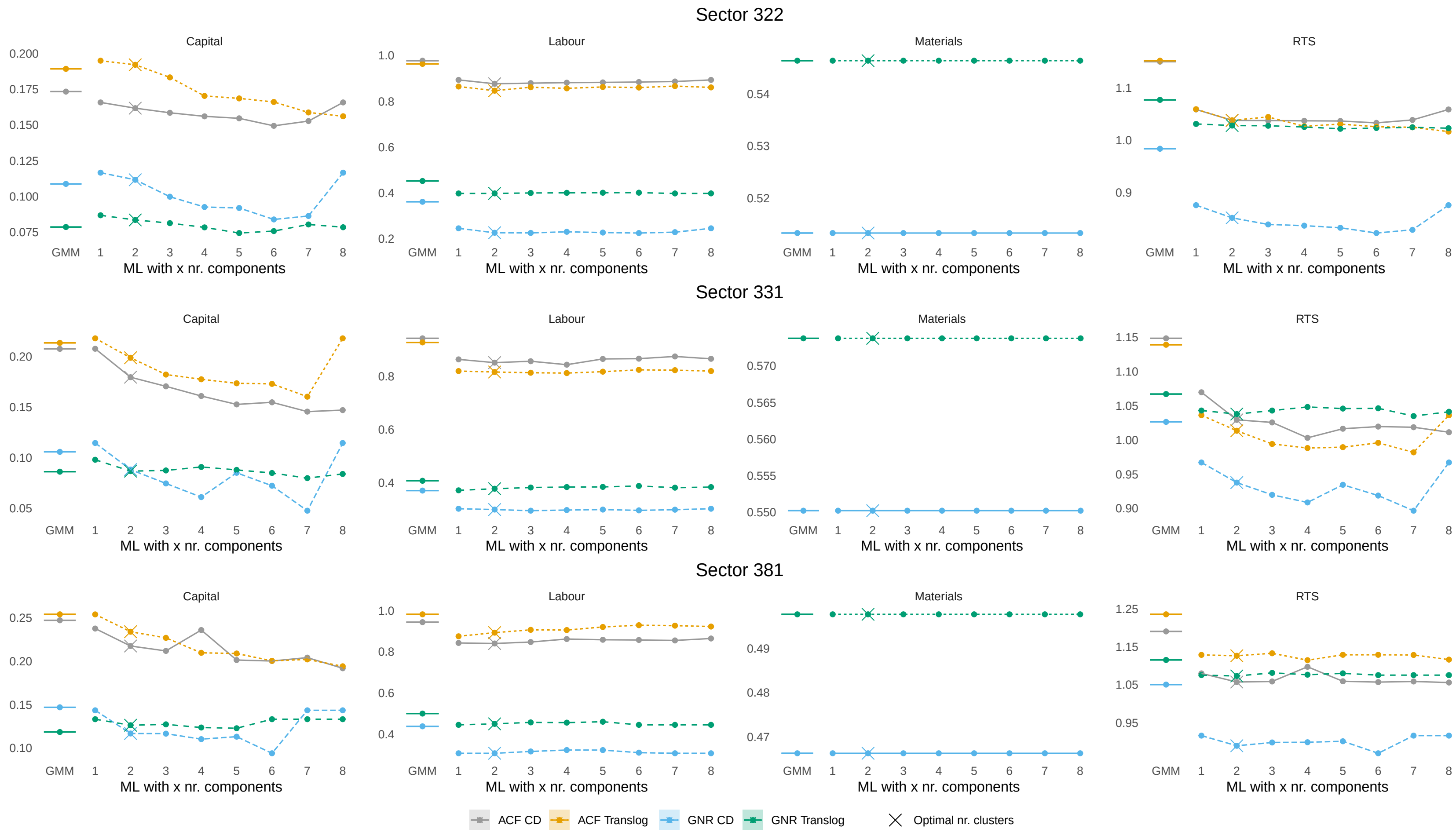

Figure D.10: Evolution of output elasticities, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with exogenous labor, in function of the number of clusters for industries 322, 331, and 381 of the Chilean economy.
Note: GMM and ML refer to the generalized method of moments and maximum likelihood estimation procedures. The times symbol indicates the optimal number of clusters as defined by the integrated complete-data likelihood Bayesian information criterion.

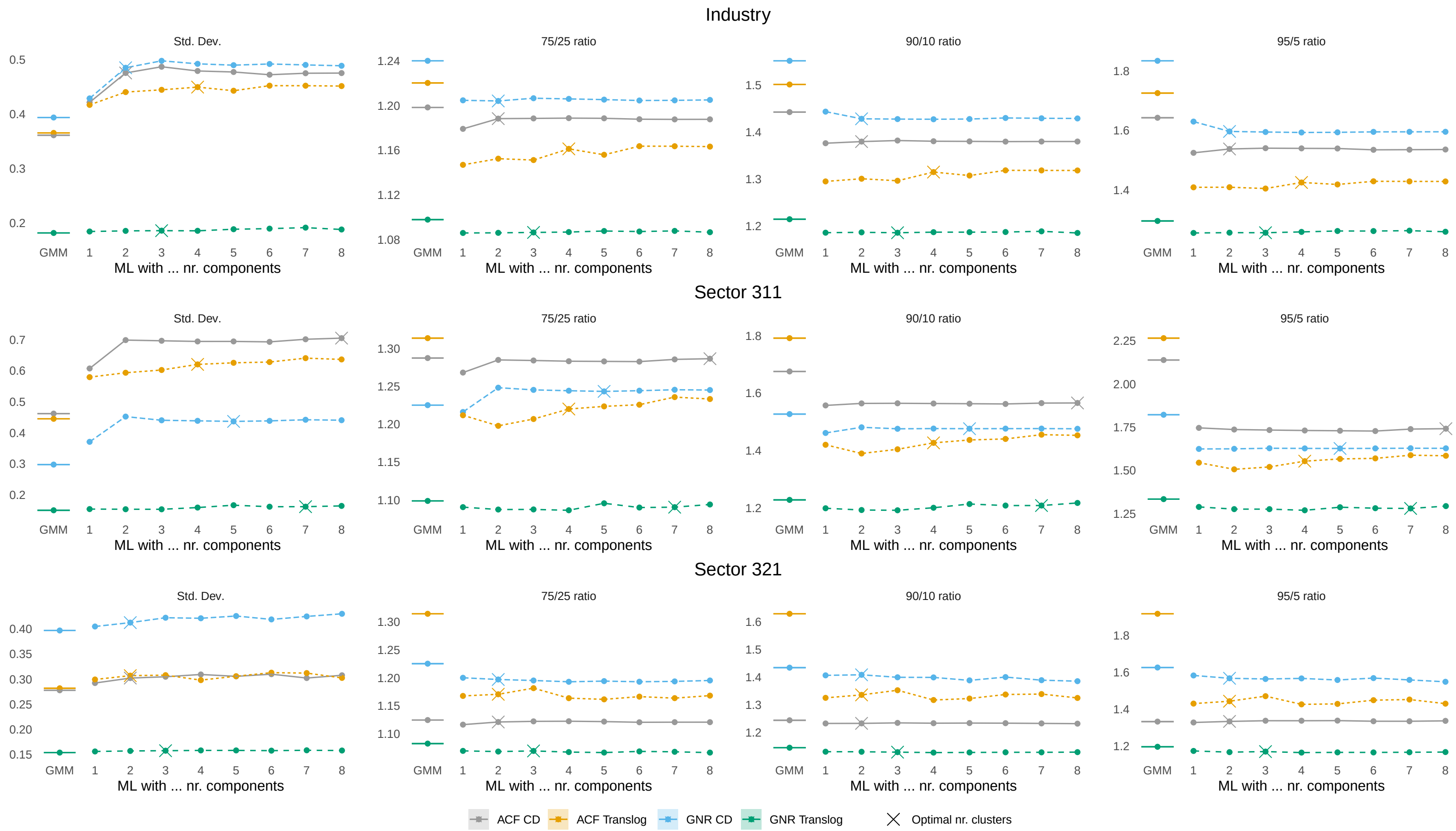

Figure D.11: Evolution of the standard deviation, 75/25-, 90/10-, and 95/5-ratio, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with exogenous labor, in function of the number of clusters for the entire manufacturing sector and industries 311 and 321 of the Chilean economy.

Note: GMM and ML refer to the generalized method of moments and maximum likelihood estimation procedures. The times symbol indicates the optimal number of clusters defined by the integrated complete-data likelihood Bayesian information criterion.

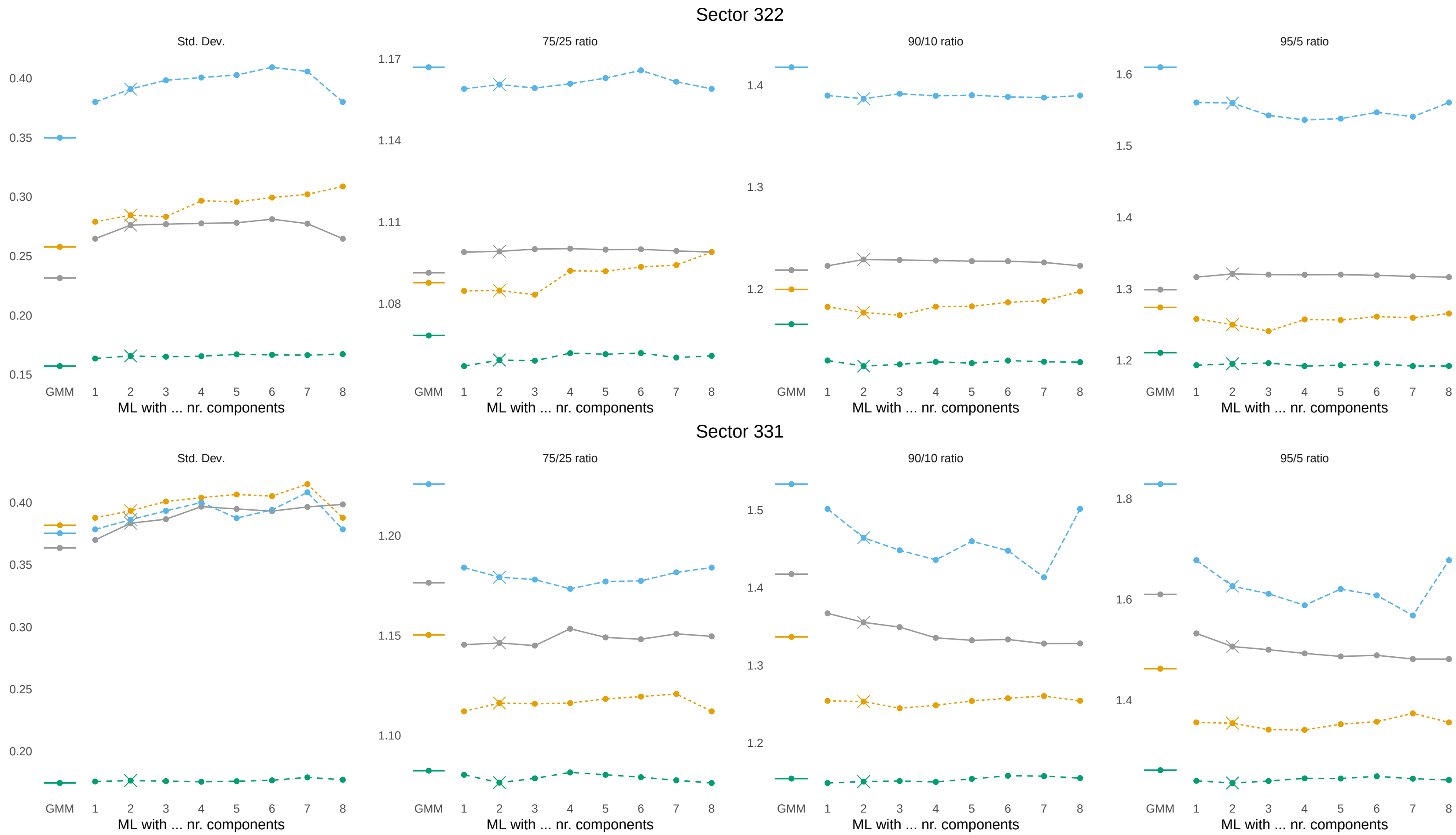

Figure D.12: Evolution of the standard deviation, 75/25-, 90/10-, and 95/5-ratio, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with exogenous labor, in function of the number of clusters for industries 321, 331, and 381 of the Chilean economy.

Note: GMM and ML refer to the generalized method of moments and maximum likelihood estimation procedures. The times symbol indicates the optimal number of clusters defined by the integrated complete-data likelihood Bayesian information criterion.

Appendix E Robustness of Production Function Coefficients

The robustness of the production function coefficients to relaxing the homogeneity assumption of the productivity growth process is not in line with existing findings in the literature (see De Loecker (2013) and the literature review in Section 2). In the robustness section of the main paper, we establish that this result is not specific to Belgian firm-level data, and here we assess the strength of this result with regard to two methodological choices made in this paper.

First, the identification strategy in this paper relies on random cluster affiliation, in contrast to the deterministic cluster affiliation currently used in the literature. Despite having demonstrated the adequacy of the random cluster affiliation identification strategy in the Monte Carlo exercise (see Section 3.5), we additionally evaluate the robustness of production function coefficients in our Belgian firm-level data using deterministic cluster identification strategies. To this end, we estimate separate production functions for 5 NACE Rev.2 industries, which are: Printing and reproduction of recorded media (18); Manufacture of rubber and plastic products (22); Manufacture of fabricated metal products, except machinery and equipment (25); Manufacture of machinery and equipment n.e.c. (28); and Manufacture of furniture (31), and an aggregate production function for the entire manufacturing sector. We parameterize the production function $f(\cdot;\boldsymbol{\beta})$ assuming both a gross-output (Gandhi et al., 2020) and value-added (Ackerberg et al., 2015) production function under both a Cobb-Douglas and Translog specification. These production functions are estimated using a GMM estimation approach with either a simple linear Markov process specification:

$$\omega_{bt} = \alpha_0 + \alpha_1 \omega_{bt-1} + \eta_{bt}, \tag{E.1}$$

or a deterministic Markov specification:

$$\omega_{bt} = \alpha_0 + \alpha_1 \omega_{bt-1} + \alpha_2 D_b + \alpha_3 \left(\omega_{bt-1} \times D_b\right) + \eta_{bt}, \tag{E.2}$$

where D_{bt} is a dummy allowing for heterogeneity in the Markov process depending on whether the firm b is respectively an exporter, importer, or engaged in FDI.

The results of this exercise, presented in Figures E.1 and E.1, confirm strong evidence of robust production function coefficients in our dataset. We observe no significant deviations between the output elasticities obtained from a linear Markov process (None) and those obtained from a Markov process allowing for heterogeneity depending on whether the firm is respectively an exporter, importer, engaged in FDI, or all three simultaneously (Export, Import, FDI, All).

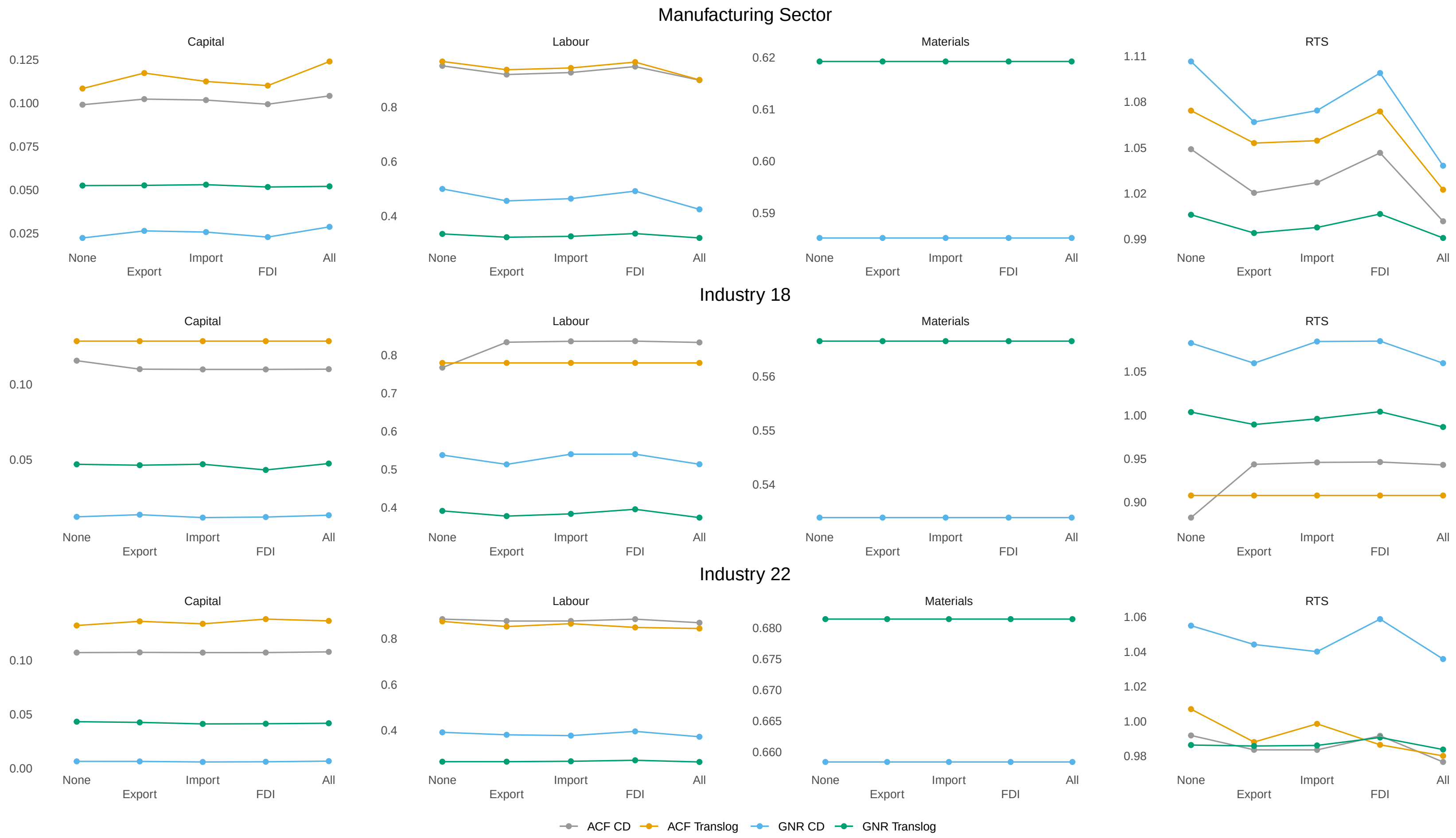

Figure E.1: Change in output elasticities, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with endogenous labor, in function of the Markov specification for the complete Industry and industries 18 and 22 of the Belgian economy.

Note: None, Export, Import, and FDI refer to a deterministic Markov specification allowing for no heterogeneity or heterogeneity in the Markov process, respectively depending on whether the firm is an exporter, importer, engaged in FDI, or all these three simultaneously.

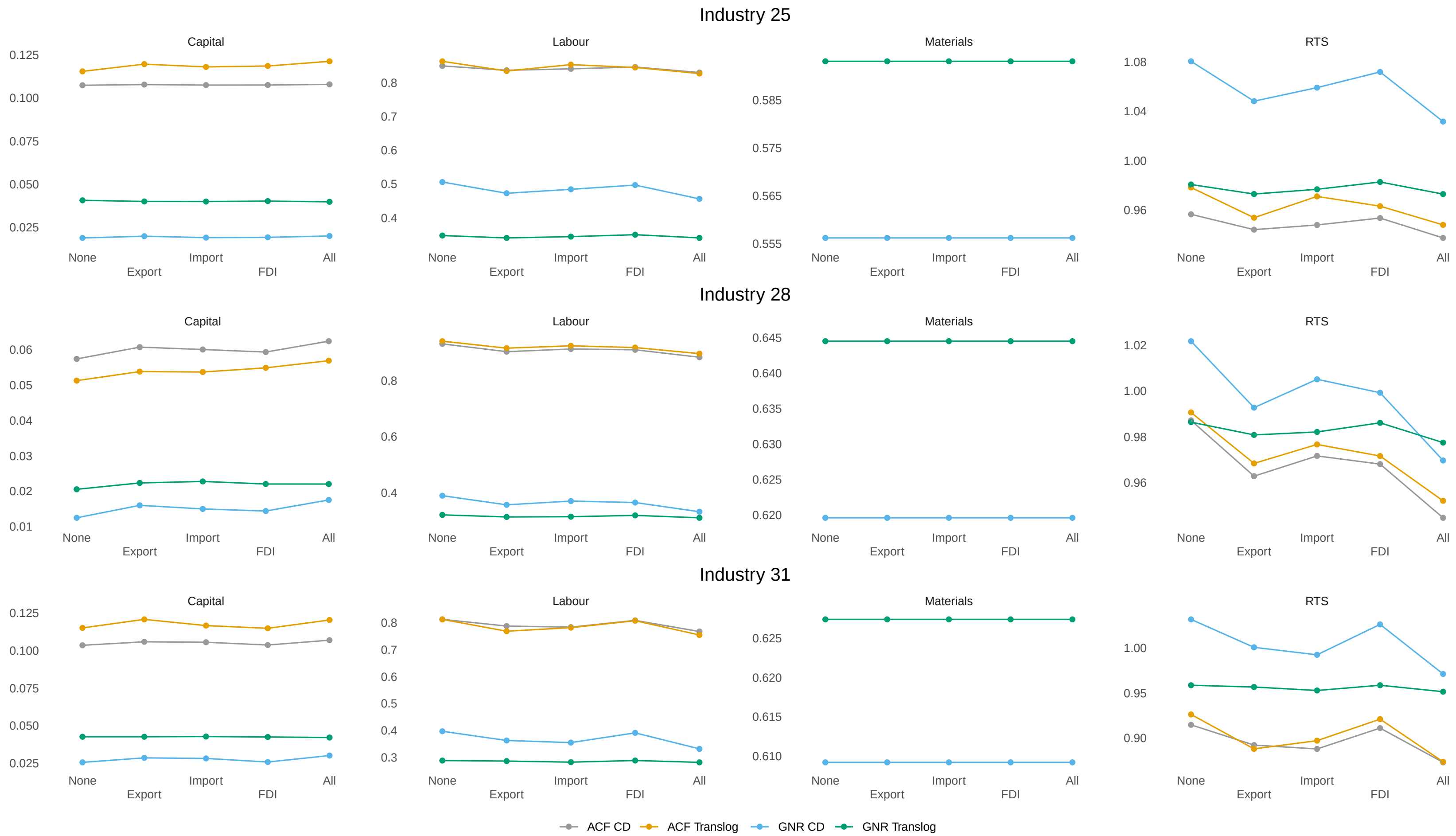

Figure E.2: Change in output elasticities, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with endogenous labor, in function of the Markov specification for the industries 25 and 28, and 31 of the Belgian economy.
Note: None, Export, Import, and FDI refer to a deterministic Markov specification allowing for no heterogeneity or heterogeneity in the Markov process, respectively depending on whether the firm is an exporter, importer, engaged in FDI, or all these three simultaneously.

Second, this paper relies on the commonly used scalar unobservability assumption, stating that materials are a flexible factor input that is decided upon simultaneously at time t without affecting future profits ($m_{bt} = h(\omega_{bt}, k_{bt}, l_{bt})$). Under the scalar unobservability assumption, the change in future productivity (and cluster affiliation) does not affect the choice of material inputs (Ackerberg, 2021). It could be, however, that a firm's cluster affiliation affects its input demand. For instance, a firm's export status has been argued to lead to differences in optimal input demand across firms (De Loecker and Warzynski, 2012). Suppose this export status is a determinant of cluster affiliation. In that case, the cluster affiliation might then also affect optimal input demand, such that $m_{bt} = h(\omega_{bt}, k_{bt}, l_{bt}, z_b^s)$. This would be in line with Kasahara et al. (2015) assuming the material input demand depends on cluster affiliation. Shenoy (2020) provides a formal framework to evaluate the adequacy of the scalar unobservability assumption. The author demonstrates that failing to account for relevant variables affecting input demand is equivalent to introducing a non-classical measurement error in the first stage of the production function estimation procedure. If this is the case, our first-stage estimation procedure might be misspecified, with unobserved heterogeneity largely being captured by the first-stage residual ε_{bt}. This could explain the robustness of our second-stage production function estimation results to the homogeneity assumption of productivity growth. However, it is unclear why one would argue that the FOC for the perfectly flexible input may be cluster-dependent while FOCs of the non-flexible inputs are not. Cluster-dependent FOCs for all inputs imply a cluster-specific production function specification, which falls outside the scope of this paper. In the concluding Section 6, we discuss the possibilities that the methodology proposed in this paper opens for future research, including cluster-dependent production function specifications.

Appendix References

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

Ackerberg, D. A. (2021). Comment on "olley and pakes-style production function estimators with firm fixed effects". Oxford Bulletin of Economics and Statistics 83(3), 836–840.

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification Properties of Recent Production Function Estimators. Econometrica 83(6), 2411–2451.

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

De Loecker, J. and F. Warzynski (2012). Markups and firm-level export status. American Economic Review 102(6), 2437–2471.

Gandhi, A., S. Navarro, and D. A. Rivers (2020a). Data archive for: On the identification of gross output production functions. Technical report, Journal of Political Economy.

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

Kasahara, H., P. Schrimpf, and M. Suzuki (2015). Identification and estimation of production function with unobserved heterogeneity. Technical report.

Kutlu, L. (2010). Battese-coelli estimator with endogenous regressors. Economics Letters 109(2), 79–81.

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

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

Shenoy, A. (2020, 03). Estimating the Production Function Under Input Market Frictions. The Review of Economics and Statistics, 1–45.