Identifying Unobserved Heterogeneity in Productivity

Ruben Dewitte, Catherine Fuss and Angelos Theodorakopoulos*

20th January 2022

Abstract

Identifying the micro-economic engines of productivity growth is a quest of general interest, but heavy data requirements hamper current identification strategies. Whereas the evolution of firm-level productivity is influenced by firm-level characteristics such as innovation, management practices, trade . . . , these characteristics are often unavailable to the researcher. Therefore, productivity measurement and the identification of its main determinants have been argued to suffer from an omitted variable bias as unobserved heterogeneity is left uncaptured. This paper proposes a methodology to uncover and, if present, correct a time-invariant omitted variable bias in the measurement of firm-level productivity. It offers an identification methodology that captures unobserved heterogeneity in productivity using Finite Mixture Models, a flexible, semi-parametric extension of state-of-the-art estimation techniques. A Monte Carlo analysis demonstrates how the proposed estimation methodology can correct the time-invariant omitted variable bias in current production function estimates. We demonstrate the applicability of the proposed method with an application to Belgian firm-level data.

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

JEL Codes: L11

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

1 Introduction

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

While identifying the micro-economic engines of productivity growth is essential for policy guidance, heavy data requirements hamper current identification strategies. Whereas the evolution of firm-level productivity is known to differ between clusters of firms driven by firm-level characteristics such as innovation (Costantini and Melitz, 2008; Bee et al., 2011; Atkeson and Burstein, 2010), management practices (Caliendo et al., 2020; Bloom and Van Reenen, 2011), trade (Kasahara and Rodrigue, 2008; De Loecker, 2013; Kasahara and Lapham, 2013), industry linkages (Luttmer, 2007) ..., these characteristics are often unavailable to the researcher. Therefore, productivity measurement and the identification of its main determinants have been argued to suffer from an omitted variable bias, since unobserved heterogeneity in the productivity growth process is left uncaptured (De Loecker, 2013).

This paper proposes a new estimation methodology to uncover and, when present, correct for the time-invariant omitted variable bias in the measurement of firm-level productivity. We build on the observation of Dewitte et al. (2020) that firm productivity is currently captured assuming a homogeneous random growth process for all firms (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015; Gandhi et al., 2020; Forlani et al., 2016). If heterogeneity from firm-level characteristics is present but not controlled for, productivity estimates might be biased, imposing heavy data requirements on current productivity estimation methodologies. This project proposes an extension to current productivity estimation methodologies using Finite Mixture Models (FMM). FMMs allow the productivity evolution to differ between time-invariant clusters of firms, even though the immediate drivers of these differences are 'unobserved'.

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

We build on a distributional assumption regarding the firm's affinity for productivity clusters to model the probability of cluster affiliation for each firm and cluster. This modelling does not require any information on the drivers of these clusters beyond the information contained in firm-level factor inputs and output. Our work then demonstrates that the production function parameters are identified based on the independence between timing decisions of factor input choices and *cluster-probability weighted* shocks to productivity (see Gandhi et al. (2020) for a recent overview of current prevalent identification schemes). To account for 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. We use the Expectation-Maximization (EM) algorithm to estimate the production function parameters and simultaneously identify a firm's cluster affiliation. Thus, by allowing for unobserved heterogeneity in productivity, we can obtain 'unbiased' estimates of productivity while reducing data requirements.[1]

We demonstrate the appropriateness of the proposed methodology in a Monte Carlo analysis and the applicability to Belgian firm-level data. First, we extend the Monte Carlo exercise from Ackerberg et al. (2015) to account for unobserved heterogeneity in productivity. We demonstrate the superiority of the proposed method relative to current estimators if firm-level drivers of unobserved heterogeneity are absent or burdened by measurement error. Second, we show that our estimator provides reasonable estimates in real-life data. In line with existing literature, we find strong evidence of heterogeneity in the evolution of productivity. In every industry we investigate, we observe multiple firm clusters that differ in their productivity growth: they differ in the level of productivity, the size of unexpected productivity shocks to productivity, and the persistence of these shocks over time. We find that cluster affiliation is positively correlated with the initial conditions of the firm, i.e., initial productivity and initial factor input use. As a result, the estimates of the proposed methodology remain stable across different estimation specifications that increasingly control for firm-level characteristics such as firm-level age, export, import, and FDI status. This performance contrasts with the dependence of commonly used productivity estimation methodologies on such firm-level characteristics. We highlight this contrast by exposing the variability in the point estimates of exporting, importing, and FDI productivity premia obtained from commonly used production function estimators, relative to the stability of these point estimates across different specifications of the proposed methodology.

The rest of the paper is structured as follows. Section 2 embeds the paper in the current scientific literature before exposing and subjecting the proposed methodology to a Monte Carlo analysis in section 3. We apply the method to Belgian firm-level data in section 4 and discuss the robustness of these results in section 5. We finish by summarizing the main contributions and future research opportunities in section 6.

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

2 Literature Review

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

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

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

The predominant specification displayed in equation 1 assumes the evolution of productivity as a function of its lagged values and a random error term: a homogeneous random growth process for all firms. Any determinative heterogeneity, as a function of specific firm-level characteristics represented by the vector $\boldsymbol{e}_{bt}$, that results in a different growth process between clusters of firms needs to be specified:

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

Such a specification allows for idiosyncratic evidence on the endogenous evolution of firm-level productivity to firm-level characteristics such as innovation (Aw et al., 2011; Doraszelski and Jaumandreu, 2013), trade (Amiti and Konings, 2007; Das et al., 2007; Kasahara and Rodrigue, 2008; De Loecker, 2013; Kasahara and Lapham, 2013), engagement in Foreign Direct Investment (FDI) (Javorcik, 2004; Blalock and Gertler, 2008), management practices (Caliendo et al., 2020; Bloom and Van Reenen, 2011), human capital Van Beveren and Vanormelingen (2014) . . . (see Table 1 for a non-exhaustive overview of the literature relying on such identification strategy).

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

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

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

Therefore, this paper proposes extending current methodologies using Finite Mixture Modeling. FMMs allow us to identify time-invariant clusters of firms with a homogeneous growth process

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

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

We are not the first that try to generalize the specification of the Markov process. Originally, Olley and Pakes (1996) envisioned a non-parametric specification of the productivity growth process, but found this to be practically infeasible (Olley and Pakes, 1996, footnote 23, p.1279). Lee et al. (2019); Gandhi et al. (2020); Ackerberg (2021) discuss the feasibility of allowing for firm-fixed effects in productivity. 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. The proposed methodology in this paper generalizes the productivity evolution semi-parametrically, allowing for cluster-specific constants that maintain sufficient information to identify the capital coefficient.

The advantages of FMMs for estimating firm-level productivity have already been explored in the productivity literature. Closest related to our work are Van Biesebroeck (2003), Kasahara et al. (2017) and Battisti et al. (2020). Van Biesebroeck (2003) specifies a non-Hicks neutral technology that differs between two technological (lean and mass) clusters of firms, with the possibility of transition from the lean-to mass technology cluster, for a panel of U.S. automobile assembly plants. The probability of cluster affiliation is modeled as a function of the relative profitability of the lean vs. mass technology in reduced form. To control for the simultaneity problem when estimating the value-added production function, Van Biesebroeck (2003) relies on a parametrization of the first-order conditions with respect to factor inputs. Similarly, Kasahara et al. (2017); Battisti et al. (2020) specify non-Hicks neutral technology with an a priori undefined number of clusters without the possibility of transition. Again, the authors rely on a parametrization of the first-order conditions with respect to factor inputs to control for simultaneity when estimating the value-added (Battisti et al., 2020) or revenue (Kasahara et al., 2017) production function. This paper focuses on generalizing prevalent Hicks-neutral productivity estimation strategies using both value-added (Ackerberg et al., 2015) and revenue (Gandhi et al., 2020) production functions. We control for simultaneity problems relying on a Limited Information Maximum Likelihood specification, preventing us from imposing additional assumptions regarding the first-order conditions of factor inputs. Moreover, we model the probability of belonging to a specific cluster according to the specified behavioral framework. This results in a Mixture-of-experts specification (Gormley and Frühwirth-Schnatter, 2019) that improves cluster identification and allows to evaluate the correlation between identified clusters and firm-level characteristics.

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

3 Methodology

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

3.1 Behavioral framework

We assume a dynamic heterogeneous firms model with cluster-dependent uncertainty in future, Hicks-neutral, productivity. The data consists of a (short) panel of firms $b = 1, \ldots, B$ over period $t = 0, \ldots, T$. These firms produce a certain output Y_{bt} provided a certain amount of capital K_{bt}, labor L_{bt} and materials M_{bt} in perfectly competitive output and input markets.[2] The firms have access to information at time t, referred to as the information set $\mathcal{I}_{bt}$. This information set $\mathcal{I}_{bt}$ is available to the firm when making its production decisions in period t. For a generic input $X_{bt} \in \{K_{bt}, L_{bt}, M_{bt}\}$, we say that this input is *non-flexible* if it is either predetermined $X_{bt} \in \mathcal{I}_{bt}$ or dynamic $X_{bt} = f\left(X_{bt-1}\right)$. Similarly, it is a *flexible* input if it is neither predetermined $X_{bt} \notin \mathcal{I}_{bt}$ nor dynamic $X_{bt} \neq f\left(X_{bt-1}\right)$ (Gandhi et al., 2020). In this paper, we assume that capital and labor are non-flexible while materials is a flexible input.

The relationship between outputs and inputs is of the form:

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

where lowercase variables indicate log values. $f^{klm}(\cdot)$ represents the production function that explains the variability in firm-level output, next to two Hicks-neutral productivity shocks. $\epsilon_{bt} \notin \mathcal{I}_{bt}$ represents an ex-post Hicks-neutral productivity shock and possible measurement

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

error that does not affect future output. $\omega_{bt} \in \mathcal{I}_{bt}$, on the other hand, represents Hicks-neutral productivity that is known to the firm before making its period t decisions. ω_{bt} "might represent variables such as the managerial ability of a firm, expected down-time due to machine breakdown, expected defect rates in a manufacturing process, soil quality, or the expected rainfall at a particular farm's location". In contrast, ϵ_{bt} "might represent deviations from expected breakdown, defect, or rainfall amounts in a given year" (Ackerberg et al., 2015, p.2414). Furthermore, the productivity component ω_{bt} evolves according to a *cluster-dependent* first-order Markov process

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

meaning that 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 evolution over time. $\mathbb{I}(\cdot)$ represents the indicator function and $p(\cdot)$ indicates the probability density function.

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

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

where $\beta \in (0,1)$ is the discount factor, $\pi_{bt}(\cdot)$ a firm's profit and $\epsilon(\cdot)$ is a choice-specific i.i.d. variable that captures the affinity of a firm for a certain cluster.[4] This affinity is observed by the firm but not by the econometrician. Integrating out the affinity provides us with the probability of belonging to a certain cluster from period 1 onwards conditional on, by the researcher observed, initial capital, labor, and productivity:

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

Equation 7 is key in acquiring an accurate identification of cluster affiliation in our empirical strategy, as we demonstrate below. Notice how the specification relates to Olley and Pakes (1996), who specify the unobserved exit policy rule of a firm as a function of observable state variables. While Olley and Pakes (1996) rely on their specification to model the probability of firm exit and control for selection bias, we will take advantage of the information available in

[3]The idea of fixed cluster membership over time is not at odds with the Belgian firm-level data we rely on in section 4. Over ten years, 100% of the Belgian firms do not change their location, 91.9% of the firms do not change their Industry affiliation, 81.8% of the firms do not change their export status, 72.6% of the firms do not change their import status, and 98.1% of the firms do not change their FDI status. Moreover, one can expect that imminent status changes of firms will reflect on their initial conditions and, therefore, result in accurately differentiated clusters.

[4]This affinity could, for instance, translate in a lower fixed cost to affiliate to a cluster.

the observed initial state variables to identify unobserved cluster affiliation.

3.2 Production function estimation

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

In the *first stage*, one tries to separate the ex-post productivity shock (ε_{bt}) and the contribution of the flexible production factors from the main estimating equation (eq. 4). Different methods exist to achieve this, without consensus on the superiority of any of these methods. For instance, Gandhi et al. (2020) build on the first-order conditions of the flexible production factor m_{bt} to identify this productivity shock and the output elasticity of the flexible production factor simultaneously. Ackerberg et al. (2015), on the other hand, assume proportionality of the flexible production factor m_{bt} to output. They then use this flexible production factor as a control variable for productivity to identify the ex-post productivity shock and measurement error component ε_{bt}.[5]

Regardless of the methodology relied upon, the first stage results in an equation of the form:

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

where ϕ_{bt} represents non-flexible output variation, i.e., what is left of output variation once the ex-post productivity shock and measurement error, and the contribution of the flexible production factor to firm-level output are subtracted.

The *second stage*, then, focuses on identifying the output elasticities of non-flexible inputs. It relies on the Markov property of productivity (see eq. 5) to replace the unobserved productivity term ω_{bt} in the previous equation (eq. 8) as a function of observables and the assumed evolution of this productivity over time. We explicitly allow for this evolution to depend on the time-invariant cluster affiliation of a firm through the cluster affiliation indicator z_b^s:[6]

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

Identification of the production function coefficients then relies on the independence of cluster-specific expected productivity deviations η_{bt}^s from the current capital stock k_{bt}, lagged non-flexible output variation ϕ_{bt-1} and, depending on the timing assumption of labor input de-

[5]See Online Appendix B for a complete workout of the first-stage for both methodologies.

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

cisions, either current l_{bt} or lagged labor l_{bt-1} : (Ackerberg et al., 2015; Arcidiacono and Jones, 2003):

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

To estimate equation 9, we rely on a Finite Mixture specification for the evolution of productivity and the Expectation-Maximization algorithm to jointly estimate the production function parameters and unobserved firm cluster affiliation. We rely on the behavioral framework presented in section 3.1. and a parametric assumption on the firm's affinity towards specific clusters to specify cluster probabilities without additional information. If there are endogeneity concerns regarding labor input decisions (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 Limited Information Maximum Likelihood specification. We start below by specifying the observed likelihood, conditioning on lagged labor values. We then determine the complete likelihood that accounts for cluster affiliation before presenting the estimation algorithm and discussing model selection algorithms over the number of clusters S.

3.2.1 Observed likelihood

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

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

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

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

[7] 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 assumed to be able to approach all distributions (McLachlan and Peel, 2000). This argument implies, however, that the number of mixture 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 (CLT): firm productivity will approximately be normally distributed if it is the sum of many independent random variables. This corresponds with 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)).

Notice 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:[8]

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

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

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

$$p^o(\boldsymbol{\phi}_b, \boldsymbol{l}_b|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b, z_b^s; \boldsymbol{\theta}^s) = \prod_{t=1}^{T} p(\phi_{bt}, l_{bt}|k_{bt}, l_{bt}, \phi_{bt-1}, l_{bt-1}, k_{bt-1}, z_b^s; \boldsymbol{\theta}^s). \tag{14}$$

To arrive at the probability of observing the complete data series, we need to model the conditional probability of belonging to a specific cluster from period one onwards (see eq. 7). We rely on the behavioral framework presented in section 3.1. and a parametric assumption on the firm's affinity towards specific clusters to specify the probabilities of unobserved cluster affiliation. We specify these probabilities as a function of the observed initial conditions, relying only on firm-level information already available for the production function estimation. Assuming the unobserved choice-specific i.i.d. variable $\epsilon(z_b^s)$ follows a type-1 extreme value distribution and relying on the reduced form of the conditional choice probability (eq. 7), we can model this probability as (McFadden, 1973):[10,11]

$$Pr(z_b^s|k_{b0}, l_{b0}, \omega_{b0}; \boldsymbol{\gamma}^1, \ldots, \boldsymbol{\gamma}^s) = \frac{e^{\gamma_0^i + \gamma_k^i k_{b0} + \gamma_l^i l_{b0} + \gamma_\omega^i \omega_{b0}}}{\sum_{s=1}^{S} e^{\gamma_0^s + \gamma_k^s k_{b0} + \gamma_l^s l_{b0} + \gamma_\omega^s \omega_{b0}}}, \qquad \forall i = 1, \ldots, S. \tag{15}$$

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

$$\begin{aligned} p^o(\boldsymbol{\phi}, \boldsymbol{l}; \underbrace{\{\boldsymbol{\gamma}^1, \ldots, \boldsymbol{\gamma}^S, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^S\}}_{=\boldsymbol{\Theta}}) = \\ \prod_{b=1}^{B} \sum_{s=1}^{S} Pr(z_b^s|k_{b0}, l_{b0}, \omega_{b0}; \boldsymbol{\gamma}^s) p^o(\boldsymbol{\phi}_b, \boldsymbol{l}_b|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b, z_b^s; \boldsymbol{\theta}^s). \end{aligned} \tag{16}$$

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

[9]In line with current productivity estimation literature, we condition on the exogenous initial output ϕ_{b0}. This means that initial output, in line with the behavioral framework set out in Section 3.1, is assumed independent from cluster membership.

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

[11]The logit specification relies on the Conditional Independence Assumption. It follows from the specification in equation 6 that this assumption is satisfied.

3.2.2 Complete log-likelihood

From equation 16, we can specify the observed log-likelihood

$$\mathcal{L}^o(\mathbf{\Theta}) = \sum_{b=1}^{B}\sum_{s=1}^{S} log\left(Pr(z_b^s|k_{b0}, l_{b0}, \omega_{b0}; \boldsymbol{\gamma}^s)p^o(\boldsymbol{\phi}_b, \boldsymbol{l}_b|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b, z_b^s; \boldsymbol{\theta}^s)\right), \tag{17}$$

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

$$\mathcal{L}^c(\mathbf{\Theta}, \boldsymbol{z}) = \sum_{b=1}^{B}\sum_{s=1}^{S} z_b^s log\left(Pr(z_b^s|k_{b0}, l_{b0}, \omega_{b0}; \boldsymbol{\gamma}^s)p^o(\boldsymbol{\phi}_b, \boldsymbol{l}_b|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b, z_b^s; \boldsymbol{\theta}^s)\right), \tag{18}$$

which forms the basis for our estimation procedure.

3.2.3 Estimation procedure

We estimate the parameters of interest based on equation 18 relying on the Expectation-Maximization algorithm (McLachlan and Peel, 2000; Miljkovic and Grün, 2016). This algorithm consists of maximizing the complete log-likelihood in an iterative procedure. Assume parameter values in iteration j are represented as $(\mathbf{\Theta})^j = \left\{(\boldsymbol{\gamma}^1)^j, \ldots, (\boldsymbol{\gamma}^S)^j, (\boldsymbol{\theta}^1)^j, \ldots, (\boldsymbol{\theta}^S)^j\right\}$.

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

$$\hat{z}_b^s = Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; (\mathbf{\Theta})^j) = \frac{Pr(z_b^s|k_{b0}, l_{b0}, \omega_{b0}; (\boldsymbol{\gamma}^s)^j)p^o(\boldsymbol{\phi}_b, \boldsymbol{l}_b|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b, z_b^s; (\boldsymbol{\theta}^s)^j)}{p^o\left(\boldsymbol{\phi}; (\mathbf{\Theta})^j\right)}. \tag{19}$$

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

(i)

$$\max_{(\boldsymbol{\theta})^{j+1}} \sum_{b=1}^{B}\sum_{s=1}^{S} Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; (\mathbf{\Theta})^j) log\left(p^o(\boldsymbol{\phi}_b, \boldsymbol{l}_b|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b, z_b^s; (\boldsymbol{\theta}^s)^{j+1});\right) \tag{20}$$

(ii)

$$\max_{(\boldsymbol{\gamma}^s)^{j+1}} \sum_{b=1}^{B}\sum_{s=1}^{S} Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; (\mathbf{\Theta})^j) log\left(Pr(z_b^s|k_{b0}, l_{b0}, \omega_{b0}; (\boldsymbol{\gamma}^s)^{j+1})\right). \tag{21}$$

This iterative process continues until there is relative stability between iteration j and $j+1$ in terms of the observed log-likelihood (eq. 17). See Online Appendix B for an in-depth 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 in light of the existence of clusters in the productivity process. We consider (i) unitary cluster affiliation and (ii) deterministic cluster affiliation based on the literature and discuss their connection to the (iii) random cluster affiliation advanced in this paper.

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

$$\phi_{bt} = f^{kl}\left(k_{bt}, l_{bt}\right) + g^1\left(\phi_{bt-1} - f^{kl}\left(k_{bt-1}, l_{bt-1}\right)\right) + \eta^1_{bt}. \tag{22}$$

While this is the specification commonly relied upon in the literature, this specification results in an omitted variable bias if there is a least one group of firms that evolves differently over time, $S > 1$:[12]

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

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

$$\hat{z}^i_b = \mathbb{I}\left(\boldsymbol{e}_b = e^i_b\right), \qquad \forall i = 1, \ldots, S, \tag{26}$$

and, as a result,

$$E\left[\sum_{s=1}^{S} \mathbb{I}\left(\boldsymbol{e}_b = e^s_b\right) \eta^s_{bt} \middle| k_{bt}, l_{bt(-1)}, \phi_{bt-1}\right] = 0. \tag{27}$$

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

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

$$\begin{aligned}\phi_{bt} = \beta_k k_{bt} + \beta_l l_{bt} &+ \mathbb{I}_b\left(s=1\right)\left(\alpha^1_0 + \alpha^1_1\left(\phi_{bt-1} - \beta_k k_{bt-1} - \beta_l l_{bt-1}\right) + \eta^1_{bt}\right) &(23)\\ &+ \mathbb{I}_b\left(s=2\right)\left(\alpha^2_0 + \alpha^2_1\left(\phi_{bt-1} - \beta_k k_{bt-1} - \beta_l l_{bt-1}\right) + \eta^2_{bt}.\right). &(24)\end{aligned}$$

When imposing unitary cluster affiliation, the specification becomes

$$\phi_{bt} = \beta_k k_{bt} + \beta_l l_{bt} + \alpha^*_0 + \alpha^*_1\left(\phi_{bt-1} - \beta_k k_{bt-1} - \beta_l l_{bt-1}\right) + \eta^*_{bt}. \tag{25}$$

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

(iii) Random cluster affiliation. This identification and estimation strategy set out in this paper approaches the unobserved cluster affiliation z_b^s as a random variable of which the probability can be determined solely based on the information already available:

$$E\left[\sum_{s=1}^{S} Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \hat{\boldsymbol{\Theta}})\eta_{bt}^s \middle| k_{bt}, l_{bt(-1)}, \phi_{bt-1}\right] = 0. \tag{28}$$

Therefore, this approach has the advantage of identifying the parameters of the production function without prior knowledge of firm cluster affiliation. Additionally, if prior information regarding cluster affiliation ($\boldsymbol{e}_b$) is available, the proposed approach is almost as good as (for $N = S$) or better than (for $N < S$) the deterministic approach.[13] Lastly, the reliance on a random specification implies this approach even allows for imperfect proxy variables (see the Monte Carlo exercise below).

3.4 Model selection

At this point, we still need to decide on the unobserved number of clusters, S. We approach this as a model selection problem, meaning that 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).

The BIC is based on penalizing the observed log likelihood function (eq. 17) proportional to the number of free parameters (np) in the model:

$$BIC(S) = -2\mathcal{L}^o(\hat{\boldsymbol{\Theta}}) + nplog(BT). \tag{29}$$

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

BIC does not take the clustering purposes for assessing S into account, regardless of the separation of the clusters. To overcome this limitation, we also consider ICLbic, which selects S so that the resulting mixture model leads to the clustering of the data with the largest evidence (Biernacki et al., 2000):

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

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

with the prior probabilities almost equal to unity

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

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

$$ICLbic(S) = -2\left(\mathcal{L}^o(\hat{\boldsymbol{\Theta}}) + \sum_{s=1}^{S}\sum_{b=1}^{B}\sum_{t=1}^{T} Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \hat{\boldsymbol{\Theta}}) log\left(Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \hat{\boldsymbol{\Theta}})\right)\right) + \frac{np}{2} log(BT). \tag{30}$$

The optimal model maximizes the ICLbic criterion over S. If the mixture components are well separated for a given S, then the term $\sum_{s=1}^{S}\sum_{b=1}^{B}\sum_{t=1}^{T} Pr(z_b^s|\cdot) log\left(Pr(z_b^s|\cdot)\right)$ tends to define a clear partition of the data set, with $Pr(z_b^s|\cdot)$ being close to 1 for one component and close to 0 for all other components. In this 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. In practice, the ICLbic appears to provide a stable and reliable estimation of S for real data sets (Celeux et al., 2018).

3.5 Monte Carlo

We rely on a Monte Carlo (MC) exercise to evaluate the estimator's performance proposed in this paper. We focus on the estimators' ability to recover unobserved heterogeneity in the productivity distribution while affirming the importance of controlling for unobserved heterogeneity in production function estimations. The setup of the Monte Carlo exercise closely mimics the setup of Ackerberg et al. (2015) that builds on Syverson (2001); Van Biesebroeck (2007). It deviates from Ackerberg et al. (2015) in the specification of the Markov process of productivity, which is assumed to differ between clusters of firms.[14] Specifically, it is assumed productivity, ω_{bt}, follows a Finite Mixture AR(1)-process

$$\omega_{bt} = \sum_{s=1}^{2} z_b^s \left[\alpha_0^s + \alpha_1^s \omega_{bt-1} + \eta_{bt}^s\right], \tag{31}$$

with 800 observations exogenously 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 $\eta_{bt}^s \sim \mathcal{N}\left(0, \sigma_\eta^s\right)$.

Firms make optimal choices of investment in the capital stock to maximize the expected (discounted) value of future profits, where there are convex capital adjustment costs the period- t capital stock 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 $e^{\omega_{b,t-i}}$, not $e^{\omega_{bt}}$) (Ackerberg et al., 2015). The production function is assumed Leontief in (and proportional to) materials, that is,

$$Y_{bt} = min\left\{K_{bt}^{\beta_k} L_{bt}^{\beta_l} e^{\omega_{bt}}, \beta_m M_{bt}\right\} e^{\epsilon_{bt}}, \tag{32}$$

where $\beta_k = 0.4, \beta_l = 0.6$, and $\beta_m = 1$. Given that the Leontief first order condition holds, valued added (log) production function to be estimated is

[14] See Online Appendix C for a complete description of the Monte Carlo exercise.

$$\frac{y_{bt}}{m_{bt}} = \beta_k k_{bt} + \beta_l l_{bt} + \omega_{bt} + \epsilon_{bt}. \tag{33}$$

We specify four different Data Generating Processes (DGPs) intending to validate the proposed methodology. The first, DGP1, assumes no difference in parameters of the productivity evolution across clusters, $\alpha_0^{1,2} = 1$, $\alpha_1^{1,2} = 0.7$ and $\sigma_\omega^{1,2} = 0.3$. This specification is equivalent to a specification without clusters and is identical to DGP1 of (Ackerberg et al., 2015). As such, the Monte Carlo analysis on GDP allows us to evaluate the appropriateness of the LIML vis-à-vis the traditional GMM specification, as well as the ability of our model selection methods to refute the two-cluster LIML in favor of the more appropriate one-cluster LIML specification. DGP2 introduces differences in the productivity evolution between clusters of firms, allowing 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. In the third DGP (DGP3), we assume cluster affiliation is observed so that the prior probability of cluster affiliation can be identified as:

$$Pr(z_b^s | k_{b0}, l_{b0}, \omega_{b0}; \boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2) = \frac{e^{\gamma_0^i + \gamma_k^i k_{b0} + \gamma_l^i l_{b0} + \gamma_\omega^i \omega_{b0} + \gamma_{cluster}^i \mathbb{I}(s=i)}}{\sum_{s=1}^{S} e^{\gamma_0^s + \gamma_k^s k_{b0} + \gamma_l^s l_{b0} + \gamma_\omega^s \omega_{b0} + \gamma_{cluster}^i \mathbb{I}(s=s)}}, \qquad \forall i = 1, \ldots, S. \tag{34}$$

DGP4 builds on DGP3 but assumes 10% of the firms are misclassified in clusters. Such specification allows us to compare the deterministic approach to cluster affiliation with our random approach.

We estimate equation 33 for all data-generating processes with the Ackerberg et al. (2015)-methodology in the first stage. The second stage of the estimation procedure, then, differentiates between the identification strategies. Following the discussion in Section 3.3, we assume (i) a unitary cluster affiliation according to (Ackerberg et al., 2015) (Uni. GMM), (ii) a deterministic cluster affiliation approach where the cluster proxy variable is observed (Det. GMM), as well as (iii) the in this paper developed estimation methodology for the second stage of the estimation with random one- and two- clusters affiliation imposed $S = 1, 2$ (1-comp LIML, 2-comp. LIML).[15]

The results of the Monte Carlo analysis are displayed in Table 2. Based on the results of DGP1, we can conclude that the LIML identification procedure allows us to estimate the production parameters accurately, just as well as the prevalent Unitary GMM methodology. Introducing multiple clusters in a single-cluster environment results in a loss of efficiency of the estimators, as can be deduced from the Determinative GMM and the 2-component LIML estimation results. DGP2 reveals an omitted variable bias for the capital coefficient and coefficients of the productivity growth process when cluster heterogeneity is present but not controlled for. The

[15]For all estimators, starting values for the estimations are set at $\beta_k = 0.3$, and $\beta_l = 0.7$.

Determinative GMM and the 2-component LIML accurately control for this heterogeneity, the former more efficiently than the latter. Note that the Determinative GMM does not identify the different variance across components, which is essential for valid inference. If cluster affiliation is known to the researcher, the 2-component LIML gains efficiency and obtains results close to the Determinative GMM. Lastly, when the researcher bases his analysis on faulty cluster affiliations, the Determinative GMM delivers biased estimates. The 2-component LIML remains robust, however. This behavior can be ascribed to the identification of cluster affiliation, which takes advantage of the information available in the initial conditions, the evolution of productivity, and the cluster affiliation indicator. Therefore, the information obtained from the cluster affiliation indicator can still be overruled based on initial capital, labor, productivity, and the productivity evolution observed in the data.

Table 2: Monte Carlo results

Methodology	β_k	β_l	α_0^1	α_1^1	σ_η^1	α_0^2	α_1^2	σ_η^2	$(\pi^1)^\dagger$	$(\pi^2)^\dagger$	BIC	ICLbic
						DGP1						
True coefficients	0.40	0.60	1.00	0.70	0.21	1.00	0.70	0.21	0.80	0.20	-	-
	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
Uni. GMM	0.40	0.60	1.00	0.70	0.21	-	-	-	1.00	1.00	-	-
	(0.02)	(0.01)	(0.03)	(0.01)	(0.00)	(-)	(-)	(-)	(0.00)	(0.00)	(-)	(-)
Det. GMM	0.40	0.60	1.00	0.70	0.21	1.02	0.70	0.21	1.00	1.00	-	-
	(0.02)	(0.01)	(0.04)	(0.01)	(0.00)	(0.06)	(0.02)	(0.00)	(0.00)	(0.00)	(-)	(-)
1-comp. LIML	0.40	0.60	1.00	0.70	0.21	-	-	-	1.00	1.00	5024.43	5024.43
	(0.02)	(0.01)	(0.03)	(0.01)	(0.00)	(-)	(-)	(-)	(0.00)	(0.00)	(211.35)	(211.35)
2-comp. LIML	0.38	0.61	1.04	0.69	0.21	1.05	0.69	0.21	64.56	35.44	5099.19	5751.62
	(0.02)	(0.01)	(0.08)	(0.02)	(0.01)	(0.08)	(0.02)	(0.01)	(10.63)	(10.63)	(212.55)	(255.54)
						DGP2						
True coefficients	0.40	0.60	1.00	0.70	0.21	0.80	0.77	0.25	0.80	0.20	-	-
	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
Uni. GMM	0.45	0.60	0.83	0.71	0.22	-	-	-	1.00	1.00	-	-
	(0.01)	(0.01)	(0.03)	(0.01)	(0.00)	(-)	(-)	(-)	(0.00)	(0.00)	(-)	(-)
Det. GMM	0.40	0.60	1.00	0.70	0.22	0.81	0.77	0.22	1.00	1.00	-	-
	(0.02)	(0.01)	(0.04)	(0.01)	(0.00)	(0.06)	(0.02)	(0.00)	(0.00)	(0.00)	(-)	(-)
1-comp. LIML	0.45	0.60	0.83	0.71	0.22	-	-	-	1.00	1.00	6010.96	6010.96
	(0.01)	(0.01)	(0.03)	(0.01)	(0.00)	(-)	(-)	(-)	(0.00)	(0.00)	(223.93)	(223.93)
2-comp. LIML	0.40	0.60	0.99	0.70	0.21	0.80	0.76	0.25	80.53	19.47	5981.90	6122.92
	(0.02)	(0.01)	(0.05)	(0.01)	(0.00)	(0.08)	(0.02)	(0.01)	(3.51)	(3.51)	(218.49)	(229.03)
						DGP3						
True coefficients	0.40	0.60	1.00	0.70	0.21	0.80	0.77	0.25	0.80	0.20	-	-
	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
Uni. GMM	0.45	0.60	0.83	0.71	0.22	-	-	-	1.00	1.00	-	-
	(0.01)	(0.01)	(0.03)	(0.01)	(0.00)	(-)	(-)	(-)	(0.00)	(0.00)	(-)	(-)
Det. GMM	0.40	0.60	1.00	0.70	0.22	0.81	0.77	0.22	1.00	1.00	-	-
	(0.02)	(0.01)	(0.04)	(0.01)	(0.00)	(0.06)	(0.02)	(0.00)	(0.00)	(0.00)	(-)	(-)
1-comp. LIML	0.45	0.60	0.83	0.71	0.22	-	-	-	1.00	1.00	6043.62	6043.62
	(0.01)	(0.01)	(0.03)	(0.01)	(0.00)	(-)	(-)	(-)	(0.00)	(0.00)	(199.86)	(199.86)
2-comp. LIML	0.40	0.60	1.01	0.70	0.21	0.81	0.77	0.25	79.84	20.16	5976.68	6040.83
	(0.02)	(0.01)	(0.04)	(0.01)	(0.00)	(0.06)	(0.02)	(0.01)	(1.47)	(1.47)	(200.86)	(202.18)
						DGP4						
True coefficients	0.40	0.60	1.00	0.70	0.21	0.80	0.77	0.25	0.80	0.20	-	-
	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
Uni. GMM	0.45	0.60	0.83	0.71	0.22	-	-	-	1.00	1.00	-	-
	(0.01)	(0.01)	(0.03)	(0.01)	(0.00)	(-)	(-)	(-)	(0.00)	(0.00)	(-)	(-)
Det. GMM	0.43	0.60	0.92	0.70	0.22	0.77	0.75	0.22	1.00	1.00	-	-
	(0.01)	(0.01)	(0.04)	(0.01)	(0.00)	(0.04)	(0.02)	(0.00)	(0.00)	(0.00)	(-)	(-)
1-comp. LIML	0.45	0.60	0.83	0.71	0.22	-	-	-	1.00	1.00	6054.97	6054.97
	(0.01)	(0.01)	(0.03)	(0.01)	(0.00)	(-)	(-)	(-)	(0.00)	(0.00)	(235.61)	(235.61)
2-comp. LIML	0.40	0.60	1.01	0.70	0.21	0.81	0.77	0.25	80.33	19.67	6004.69	6076.77
	(0.02)	(0.01)	(0.04)	(0.01)	(0.00)	(0.06)	(0.02)	(0.01)	(1.73)	(1.73)	(232.66)	(238.81)

Notes: Results display the average and the standard deviation (between brackets) of the coefficient estimates obtained across 100 iterations. † $\pi^i = \sum_{b=1}^{B} \frac{\mathbb{I}_b(s=i)}{B}$

4 Application to Belgian firm-level data

Having established that our estimator performs well in Monte Carlo simulations, we evaluate the performance of the proposed estimator on a database of Belgian manufacturing firms' balance sheets over the period 2008-2018. We retain a cleaned set of active firms that report output, capital stock at the beginning of the year, number of employees in FTE, and material costs.[16] This database is combined with a set of firm-level characteristics considered relevant for productivity growth, including firm age, firm sector affiliation, and whether or not the firm is part of or participates in export, import, and FDI.[17]

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 industry. 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) Cobb-Douglas and Translog specification. These production functions are estimated using either a traditional 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 (nr. clusters S, limited to $S = 10$) in a linear Markov process $g^s(\omega_{bt-1}, \boldsymbol{\alpha}^s)$. For sparsity reasons, we will discuss the estimation results for a value-added Translog production function of sector 22 and refer to the robustness section (Section 5) for a discussion of the estimation results for the auxiliary specifications.

4.1 Production function estimates

We display the production function estimation results in Table 3. The average output elasticities and Returns to Scale (RTS) shown in the first three rows of the table indicate there exists small, though not significant, differences between the GMM and 1-component LIML estimation methodology. Most likely, these differences can be ascribed to differences in efficiency. The LIML estimator is a more efficient estimator than the GMM estimator. Its instruments are constructed optimally using the nonlinear model specification $E\left[\frac{\partial \phi_{bt}}{\partial \boldsymbol{\beta}} \eta_{bt}\right] = 0$. In comparison, the GMM estimator typically relies on factor input and output levels to specifiy moment conditions (Ackerberg et al., 2015; Gandhi et al., 2020).[18]

We find little evidence of an omitted variable bias in the production function parameters when

[16]We clean the data both on levels, ratios, and on growth rates of these ratios to prevent effects of extreme outliers and extreme noise on the analysis. Specifically, we limit the sample to observations with more than one employee in FTE, deflated sales, materials, and capital to values larger than 1,000 euro, export and import ratios to the value of one, and remove firms belonging to industry NACE Rev. 19 (Coke and refined petroleum products) from the sample. Further, we removed the lowest and highest percentile of the log of the labor-, capital-, and materials-output ratio and observations with absolute growth rates of these ratios larger than 1000%.

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

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

Table 3: Estimation results for a value-added translog production function estimated with LIML for sector 22.

	GMM	LIML					
Description		1-comp.	2-comp.	3-comp.	4-comp.	5-comp.	6-comp.
Capital	0.130	0.118	0.124	0.124	0.124	0.124	0.124
	(0.016)	(0.017)	(0.017)	(0.018)	(0.016)	(0.020)	(0.020)
Labor	0.879	0.860	0.854	0.866	0.860	0.852	0.857
	(0.020)	(0.045)	(0.023)	(0.027)	(0.028)	(0.027)	(0.029)
RTS	1.009	0.978	0.979	0.990	0.984	0.977	0.981
	(0.015)	(0.038)	(0.018)	(0.020)	(0.023)	(0.026)	(0.025)
Std. Dev.	0.180	0.159	0.156	0.157	0.156	0.155	0.156
	(0.016)	(0.030)	(0.019)	(0.017)	(0.018)	(0.019)	(0.016)
75/25 ratio	1.015	1.014	1.014	1.014	1.013	1.014	1.014
	(0.002)	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
95/5 ratio	1.031	1.028	1.028	1.028	1.028	1.028	1.028
	(0.004)	(0.007)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
90/10 ratio	1.042	1.037	1.036	1.038	1.037	1.037	1.037
	(0.005)	(0.009)	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)
Nr. parameters[d]	7	20	37	54	71	88	105
NLL		-6835	-8735	-9134	-9321	-9528	-9647
BIC		-13505	-17166	-17824	-18060	-18335	-18434
ICLbic		-13505	-17121	-17714	-17903	-18163	-18246

Notes: a. Standard errors displayed between brackets are obtained from wild bootstrap clustered at the firm level with 49 replications.
b. The first three rows display the average elasticities across firms. The fourth row displays the standard deviation of the productivity estimates. The subsequent three rows report ratios of productivity for firms at various percentiles of the productivity distribution. NLL stands for Negative Log-Likelihood, BIC for the Bayesian Information Criterion and ICLbic for the Integrated Complete-data Likelihood Bayesian Information Criterion.
c. Estimates obtained from Belgian NACE Rev. 22 firm-level database over the years 2008–2018 with 4,399 observations from 626 firms.
d. Nr. of parameters refers to the number of parameters relied on in the second stage of the estimation procedure.

evaluating the evolution of the output elasticities when allowing for increasing heterogeneity in productivity growth. While point estimates evolve as the number of clusters increases, they do not significantly differ from the 1-component LIML estimates. We demonstrate in Online Appendix E that this absence of evidence in favor of an omitted variable bias for the production function parameters is not specific to the methodology proposed in this paper. It is likely related to the adequacy of the scalar unobservability assumption in current production function estimation procedures (Shenoy, 2020).

As the increasing number of clusters has a minor influence on the production function coefficients, the shape of the productivity distribution is also not significantly affected. We report the standard deviation of the productivity estimates in row four of Table 3 and ratios of productivity for firms at various percentiles of the productivity distribution in the subsequent three rows. We observe that these ratios do not significantly change with increasing clusters. The limited variation in the shape of productivity distribution can also visually be observed in Online Appendix Figures 1 and 2 where we plot the productivity densities with an increasing number of clusters.

However, this absence of an omitted variable bias is not due to the lack of heterogeneity in productivity growth. The goodness-of-fit indicators reported in the bottom three rows of the table demonstrate an increasingly good model fit as the number of clusters increases, with an optimal number of four clusters as indicated by the ICLbic decreasing up to the 6-component LIML. These 6 clusters are well-identified, as indicated by the posterior probabilities displayed in Online Appendix Figure 3.

We further investigate the heterogeneity in the evolution of productivity by displaying the cluster-specific coefficients of the productivity evolution in Table 4. We identify firm cluster affiliation by maximizing the posterior cluster affiliation probability. We observe heterogeneity in both the constant and auto-regressive parameters of the productivity process across clusters, leading to a hierarchy in the stationary average productivity levels (μ_ω) among clusters. Cluster 2, for instance, holds a clear productivity advantage over cluster 4, with a productivity premium of $\pm 20\%$. This distinction can also be visually evaluated based on the cluster-specific productivity densities displayed in Online Appendix Figure 4. Additionally, we observe significant heterogeneity in the volatility of the distribution of unexpected shocks to productivity (σ_η) and stationary volatility (σ_ω) that correlate with stationary average productivity levels. Highly volatile productivity processes correlate with a relatively higher average productivity level. This indicates that firms that end up in the right tail of the productivity distribution have done so through a relatively volatile productivity growth process. However, high volatility in productivity does not mean that this volatility is equally persistent. Using an Impulse Response Function in Figure 1, we demonstrate that an unexpected impulse to productivity has a relatively more minor long-lasting influence in a cluster with a volatile productivity growth process than a relatively stable growth process.

Table 4: Cluster-specific characterization of the productivity evolution and its stationary distribution for sector 22.

Cluster description	Prop. (%)	α_0	α_1	σ_η	μ_ω	σ_ω	$\bar{\omega}$	s_ω
Cluster 1	24.973	0.677	0.946	0.041	12.522	0.126	12.552	0.143
	(2.685)	(0.147)	(0.010)	(0.005)	(0.673)	(0.020)	(0.671)	(0.022)
Cluster 2	20.824	1.029	0.918	0.059	12.575	0.149	12.597	0.170
	(3.612)	(0.120)	(0.011)	(0.008)	(0.683)	(0.015)	(0.671)	(0.017)
Cluster 3	20.000	0.679	0.945	0.022	12.400	0.068	12.391	0.062
	(4.788)	(0.117)	(0.010)	(0.004)	(0.676)	(0.015)	(0.680)	(0.016)
Cluster 4	18.626	0.805	0.935	0.018	12.367	0.050	12.384	0.075
	(2.108)	(0.215)	(0.018)	(0.002)	(0.679)	(0.009)	(0.681)	(0.008)
Cluster 5	9.808	2.424	0.805	0.032	12.400	0.054	12.392	0.058
	(2.272)	(0.481)	(0.039)	(0.004)	(0.681)	(0.008)	(0.679)	(0.009)
Cluster 6	5.769	3.532	0.719	0.132	12.568	0.190	12.588	0.207
	(0.556)	(0.610)	(0.039)	(0.015)	(0.668)	(0.027)	(0.665)	(0.033)

Notes: a. Standard errors displayed between brackets are obtained from a clustered wild bootstrap with 49 replications.
b. Cluster affiliation determined by maximal posterior cluster affiliation probability.
c. Prop. stands for the percentage of firms affiliated to each cluster.
d. $\mu_\omega = \frac{\alpha_0}{1-\alpha_1}$, $\sigma_\omega = \sqrt{\frac{\sigma_\eta^2}{1-\alpha_1^2}}$, $\bar{\omega} = \sum_{b=1}^{B}\sum_{t=1}^{T}\frac{\omega_{bt}}{BT}$, $s_\omega = \sqrt{\frac{1}{BT}\sum_{t=1}^{T}\sum_{b=1}^{B}(\omega_{bt}-\bar{\omega_{bt}})^2}$.

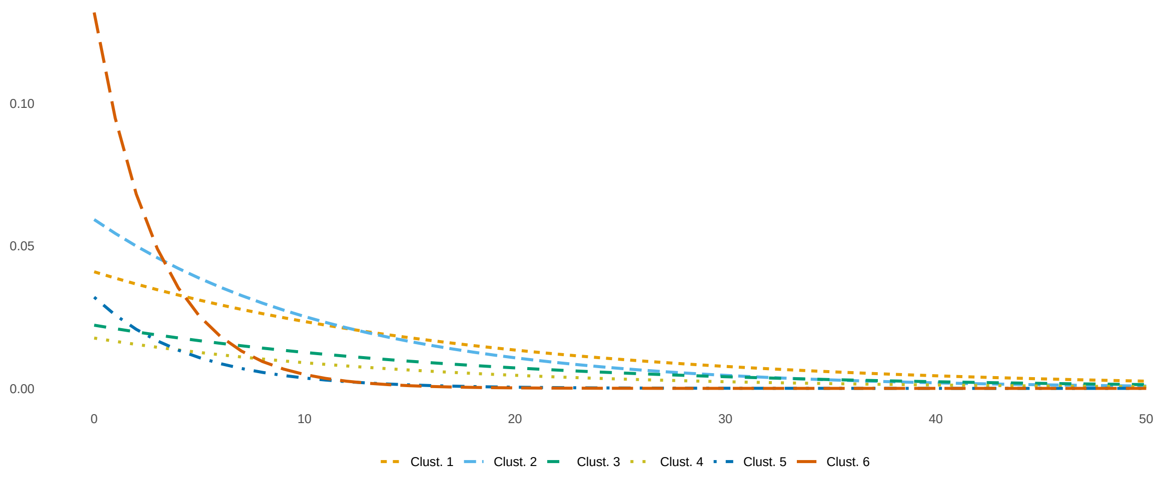

Figure 1: Impulse-Response Function of the cluster-specific productivity process for sector 22.
Note: 50-year response function of the evolution of productivity after a one standard deviation unexpected productivity impulse: $IRF^s = \sum_{t=0}^{T=50}\sigma_\eta^s(\rho^s)^t$. Cluster affiliation is determined as the maximal posterior cluster affiliation probability.

4.2 Characterizing unobserved heterogeneity in productivity

Up till now, our analysis has relied on the minimal information required to estimate a production function: factor input and output information. We have, however, additional information regarding the age and internationalization status (whether or not a firm exports, imports, and/or engages in Foreign Direct Investment (FDI)) of the firm. We exploit this additional information to highlight the advantages of the proposed estimator. First, we demonstrate the robustness of the proposed methodology to the inclusion of additional information, in contrast to the the commonly used productivity estimation techniques. Second, we showcase the importance of this robustness and accounting for unobserved heterogeneity in productivity when comparing productivity between groups of firms that differ in specific firm-level characteristics.

The estimation results reported above rely on a base specification of cluster probabilities, conditioning only on initial capital, labor, and productivity (see eq. 15). This specification originated under the assumption that the initial conditions contain sufficient information to identify cluster affiliation. If this assumption does not hold, augmenting the base specification with additional, economically relevant (see Section 2) firm-level characteristics could improve the identification of cluster affiliation. To evaluate this hypothesis, we specify an augmented multinomial regression:

$$\begin{aligned}\frac{Pr(z_b^i|\ldots;\boldsymbol{\gamma}^i)}{Pr(z_b^1|\ldots;\boldsymbol{\gamma}^1)} = \gamma_0^i &+ \gamma_1^i k_{b0} + \gamma_2^i l_{b0} + \gamma_3^i \omega_{b0} + \gamma_4^i age_{b0} \\ &+ \gamma_5^i ExportStatus_b + \gamma_6^i ImportStatus_b \\ &+ \gamma_7^i FDIStatus_b, \qquad \forall i = 2,\ldots,S\end{aligned} \tag{35}$$

where cluster probabilities are specified conditional on initial capital, labor and productivity as well as initial age, the Export, Import and FDI status over the sample period.[19] Furthermore, we also specify a version of equation 35 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 equally well as our base specification.

We re-estimate the production function relying on the augmented multinomial regression and augmented regression without initial conditions, as specified above, and report the resulting Log-likelihood, BIC, and ICLbic in Table 5. Focusing on the difference between the base specification and augmented specification first, we can conclude that the base specification 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 different 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 unobserved heterogeneity and the initial conditions. Once initial conditions are controlled for, additional firm-level characteristics have

[19] We identify firms as Inactive or Active depending on whether they are respectively inactive over the entire sample period or active over at one point in time over the sample period. Inactive firms are chosen as the reference category.

limited explanatory power. To demonstrate how instrumental initial conditions are to 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. In conclusion, even when firm-level information regarding age and the internationalization status of a firm is available, a significant part of the heterogeneity in productivity remains unobserved.

Table 5: Goodness-of-fit indicators for estimation with varying concomitant specifications in sector 22

Specification	Log-likelihood	BIC	ICLbic
Base specification	9,647.45	-18,433.93	-18,245.55
Augmented specification	9,658.52	-18,292.07	-18,112.65
Augmented specification without initial conditions	9,602.63	-18,262.30	-18,063.17

Notes: a. The base specification refers to eq. 15, the augmented specification refers to eq. 35, and the specification without initial capital and labor refers to eq. 35 without initial capital and labor.
b. BIC stands for the Bayesian Information Criterion and ICLbic for the Integrated Complete-data Likelihood Bayesian Information Criterion.
c. Estimates obtained from a Value-added Translog production function with endogenous labor estimated on the Belgian NACE Rev. 22 firm-level database over the years 2008–2018 with 4,399 observations from 626 firms.

A closer analysis of the correlation between the firm characteristics and firm cluster affiliation can be obtained from the summary statistics across firm clusters provided in Table 6. We observe that initial productivity correlates strongly with the stationary productivity levels of 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 resonates with the findings of Sterk et al. (2021), who find that initial conditions strongly determine the heterogeneity in productivity. Firm age, then, seems to correlate with the persistence of the productivity growth process. The clusters with relatively less-persistent growth processes (clusters 2, 5, and 6) hold, on average, younger firms while the clusters with a relatively persistent growth process (clusters 1, 3, and 4) contain the older firms on average. Additionally, we can observe that clusters 1 and 4, clusters linked to a relatively persistent productivity growth process, correlate positively with firm size in terms of initial output, capital, and labor, and with the internationalization status of the firm: export, import, and FDI status. Importers also have a relatively large probability of belonging to cluster 2, a relatively more volatile cluster. Provided that our productivity measures contain both efficiency and demand, we do not engage in a more detailed analysis of these heterogeneous effects.

An intriguing observation from 6, though, is that the internationalization status of firms is connected to multiple clusters. In particular, it appears that low-productivity firms that are active in export, import, and/or FDI primordially belong to cluster 4, while higher-productivity firms with an international connection primarily belong to cluster 1 (see also Online Appendix Figure 1 for a visual representation of firm cluster affiliation as a function of firm-level productivity and the internationalization status). This observation points to heterogeneity in productivity beyond what can be captured by a simple dummy variable; a common strategy relied upon in the literature (see Section 2).

We evaluate the economic repercussions of this heterogeneity by evaluating differences in the

Table 6: Average cluster characteristics for sector 22

	Overall	**Clust. 1**	**Clust. 2**	**Clust. 3**	**Clust. 4**	**Clust. 5**	**Clust. 6**
Cluster proportions (%)	100.00	24.97	20.82	20.00	18.63	9.81	5.77
log(Initial output)	15.17	16.05	14.81	14.65	16.10	13.67	14.78
log(Inital capital)	13.28	13.86	12.86	12.91	14.12	12.42	12.89
log(Initial labour)	2.78	3.37	2.05	2.67	3.97	1.80	2.03
log(Initial productivity)	12.49	12.56	12.62	12.39	12.38	12.38	12.61
Inital age	24.80	26.59	21.18	26.13	29.41	20.25	22.10
Exporter prop. (%)	65.13	78.87	59.83	51.79	81.63	46.15	60.78
Importer prop. (%)	80.68	91.55	86.32	69.64	90.82	58.46	70.59
FDI prop. (%)	10.26	14.79	3.42	7.14	21.43	3.08	7.84

Notes: Cluster affiliation determined by maximal posterior cluster affiliation probability.

productivity distribution by internationalization status when this productivity distribution is obtained under different specifications of heterogeneity in productivity. We estimate productivity using the GMM and LIML identification strategy with (i) a unitary specification of the Markov process, $\omega_{bt} = \alpha_0 + \alpha_1 \omega_{bt-1} + \eta_{bt}$, (ii) a deterministic control for the internationalization status, $\omega_{bt} = \alpha_0 + \alpha_2 IntStatus_b + \alpha_1 \omega_{bt-1} + \alpha_3 \omega_{bt-1} IntStatus_b + \eta_{bt}$, where the dummy variable $IntStatus_{b0}$ indicates the time-invariant Export, Import, or FDI status of a firm, and (iii) an exhaustive control for heterogeneity in productivity:

$$\begin{aligned} \omega_{bt} =& \alpha_0 + \alpha_1 \omega_{bt-1} + \alpha_3 Age_{b0} + \\ & \alpha_4 ExportStatus_b + \alpha_5 \omega_{bt-1} ExportStatus_b + \\ & \alpha_6 ImportStatus_b + \alpha_7 \omega_{bt-1} ImportStatus_b + \\ & \alpha_8 FDIStatus_b + \alpha_9 \omega_{bt-1} FDIStatus_b + \eta_{bt}. \end{aligned} \tag{36}$$

Similarly, we obtain productivity from the Finite Mixture LIML identification strategy with the optimal nr. of 6 clusters and (i) the base specification for cluster affiliation (see eq. 15) (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 (see eq. 35).

We display in Table 7 the average, median, and Standard Deviation productivity premia by internationalization status and estimation methodology. We can observe that both for the GMM and LIML estimation method, the premia are dependent on the underlying specification of the Markov process. The exporters' premia are generally increasing as we increasingly allow for heterogeneity in the Markov process. Point estimates for the mean, even though imprecisely estimated, increase from -0.76 % for the unitary specification over 0.28% for the deterministic specification to 0.76% for the Exhaustive Markov specification estimated with GMM. This increase can be explained by the fact that we increasingly allow for heterogeneity in the evolution of productivity of generally highly productive firms (exporters, importers, and firms active in FDI) while maintaining the linear specification for the productivity evolution of the lower-productive reference category. However, the exporters' premia obtained under the Finite Mixture LIML specification remains relatively stable across different specifications of the cluster probabilities, with point estimates that are generally lower than those obtained from the single cluster LIML specification. A similar picture can be observed for the productivity premia of Importers and firms engaged in FDI. Whereas current productivity estimation strategies rely on explicitly controlling for heterogeneity in productivity which renders inference dependent on the heterogeneity information available to the researcher, our proposed estimation strategy remains robust when confronted with unobserved heterogeneity in productivity.

Table 7: Summary statistics' productivity premia (in %) by internationalization status and estimation methodology for sector 22

Methodology	Control	Mean	Median	St. Dev.
		Exporters		
GMM	Unitary	-1.493 (0.836)	-0.338 (0.859)	-11.000 (5.540)
	Deterministic	0.276 (1.058)	0.822 (0.952)	-6.512 (9.209)
	Exhaustive	0.760 (2.388)	1.263 (2.117)	4.456 (11.759)
LIML	Unitary	2.559 (5.960)	3.088 (6.049)	8.988 (10.386)
	Deterministic	3.857 (8.824)	4.119 (9.075)	9.114 (11.959)
	Exhaustive	2.854 (5.465)	3.273 (5.941)	9.399 (12.275)
Finite Mixture LIML	Unitary	2.063 (2.108)	2.033 (1.642)	5.707 (11.813)
	Deterministic	2.092 (2.153)	2.082 (1.756)	5.747 (13.549)
	Exhaustive	2.011 (2.116)	2.007 (1.664)	5.605 (11.900)
		Importers		
GMM	Unitary	2.855 (1.416)	2.314 (1.525)	6.180 (8.120)
	Deterministic	3.918 (2.018)	3.541 (2.060)	8.786 (9.614)
	Exhaustive	5.532 (2.305)	5.095 (2.647)	26.825 (12.907)
LIML	Unitary	7.356 (6.817)	6.640 (6.958)	34.099 (12.964)
	Deterministic	8.768 (9.794)	7.956 (10.164)	34.093 (15.545)
	Exhaustive	7.753 (6.135)	6.859 (6.456)	34.632 (16.476)
Finite Mixture LIML	Unitary	6.483 (2.092)	5.777 (1.775)	29.044 (14.777)
	Deterministic	6.474 (1.570)	5.786 (1.438)	29.014 (13.239)
	Exhaustive	6.425 (2.018)	5.753 (1.650)	28.921 (15.001)
		FDI		
GMM	Unitary	3.296 (1.581)	6.645 (1.988)	-26.635 (12.004)
	Deterministic	-9.561 (16.601)	-8.505 (16.934)	16.705 (28.503)
	Exhaustive	-5.272 (16.123)	-4.802 (17.507)	11.148 (23.546)
LIML	Unitary	-3.955 (12.185)	-3.567 (12.021)	18.332 (18.377)
	Deterministic	-8.037 (4.016)	-7.165 (2.611)	24.430 (14.396)
	Exhaustive	-3.814 (11.129)	-3.269 (10.226)	19.337 (16.650)
Finite Mixture LIML	Unitary	-1.777 (3.018)	-1.824 (3.200)	6.348 (16.638)
	Deterministic	-1.784 (3.554)	-1.832 (3.712)	6.202 (17.896)
	Exhaustive	-1.821 (2.872)	-1.866 (3.015)	6.195 (17.047)

Notes: a. Standard errors displayed between brackets are obtained from wild bootstrap clustered at the firm level with 49 replications.
b. Unitary refers to a specification without additional heterogeneity in productivity, Deterministic refers to a deterministic specification of additional heterogeneity in productivity, and Exhaustive refers to an exhaustive specification of additional heterogeneity in productivity.

5 Robustness

The main text reports the estimation results for a value-added Translog production function for sector 22. We demonstrate in Online Appendix D that the reported results are robust to estimation methodology and sector selection. We evaluate the results for four alternative estimation methodologies, assuming both a gross-output (Gandhi et al., 2020) and value-added (Ackerberg et al., 2015) Cobb-Douglas and Translog specification, and all five sectors considered. The proposed method delivers reasonable production function estimates in all cases and confirms the results presented in the main text.

One could also worry that our results are specific to the Belgian firm-level dataset. We, therefore, extend our analysis to the Chilean firm-level dataset provided by (Gandhi et al., 2020). The results (see Online Appendix D) confirm the results obtained based on the Belgian firm-level dataset. In accordance with the main results, we find little evidence for a significant omitted variable bias but strong evidence favoring multiple clusters in the productivity evolution.

6 Conclusion

This paper proposes a general extension of state-of-the art production function estimation procedures to control for, and identify, unobserved heterogeneity in the evolution of productivity. We demonstrate the applicability of this methodology by means of a Monte Carlo exercise and an application to Belgian firm-level data. We provide strong evidence of heterogeneity in the evolution of productivity. This unobserved heterogeneity is positively correlated with the initial conditions of a firm, especially with initial productivity. Additional explanatory variables such as the export, import, and FDI status of a firm are correlated with multiple clusters, indicating the existence of heterogeneity in productivity beyond what is captured by these observed firm-level characteristics. As a result, current productivity estimation methodologies are dependent on the availability of such firm-level characteristics, while the proposed methodology maintains its performance in the face of supplementary information.

The methodology proposed in this paper opens exciting new avenues for research. It is of interest to every applied researcher who is interested in accurately recovering the effect of a firm-specific event (for instance, engaging in export) on the firm-level productivity evolution by allowing for unobserved heterogeneity in productivity. Furthermore, while the proposed methodology allows to correctly identify unobserved heterogeneity in productivity, questions regarding the drivers of this heterogeneity remain. Building on the newly developed productivity estimation strategy, one can systematically search for the main determinants of productivity growth. Obtaining such insights 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 for 'dissimilar' firms to grow at a different pace. The advantage of this approach is the enforcement and use of a cluster structure. Current methods work with a predefined cluster of firms (such as sector-specific productivity growth processes) and aim to find within-cluster determinants of productivity growth. The approach defined here allows the data to determine firm clusters and tries to identify the *between-cluster* determinants of productivity growth. Applying our methodology to a dataset that allows to differentiate between heterogeneity in revenue and output productivity could be very informative in this respect.

One could also extend the current estimation methodology to relax its underlying assumptions. The current estimation strategy holds a firm's cluster affiliation constant over time. In the spirit of (Van Biesebroeck, 2003), one could extend the current methodology to allow for regime switching in the evolution of productivity for economies or sectors undergoing a structural transformation. Lastly, one could extend the framework to allow for non-hicks neutral productivity by allowing for a cluster-dependent production function, similar to the work of Kasahara et al. (2017) and Battisti et al. (2020). Allowing for a cluster-dependent production function would allow for a more extensive discussion on the adequacy of the scalar unobservability assumption (Shenoy, 2020) and the size of the omitted variable bias.

References

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. 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.

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

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.

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., 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.

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.

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. *Handbook of mixture analysis*, 121–160.

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.

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

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.

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

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

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.

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.

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. 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 (2017). Identification and estimation of production function with unobserved heterogeneity. *mimeo*.

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.

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.

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.

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

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

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.

Online Appendix to "Identifying Unobserved Heterogeneity in Productivity"

Ruben Dewitte, Catherine Fuss and Angelos Theodorakopoulos*

20th January 2022

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

Appendix A Additional Figures and Tables

A.1 Figures

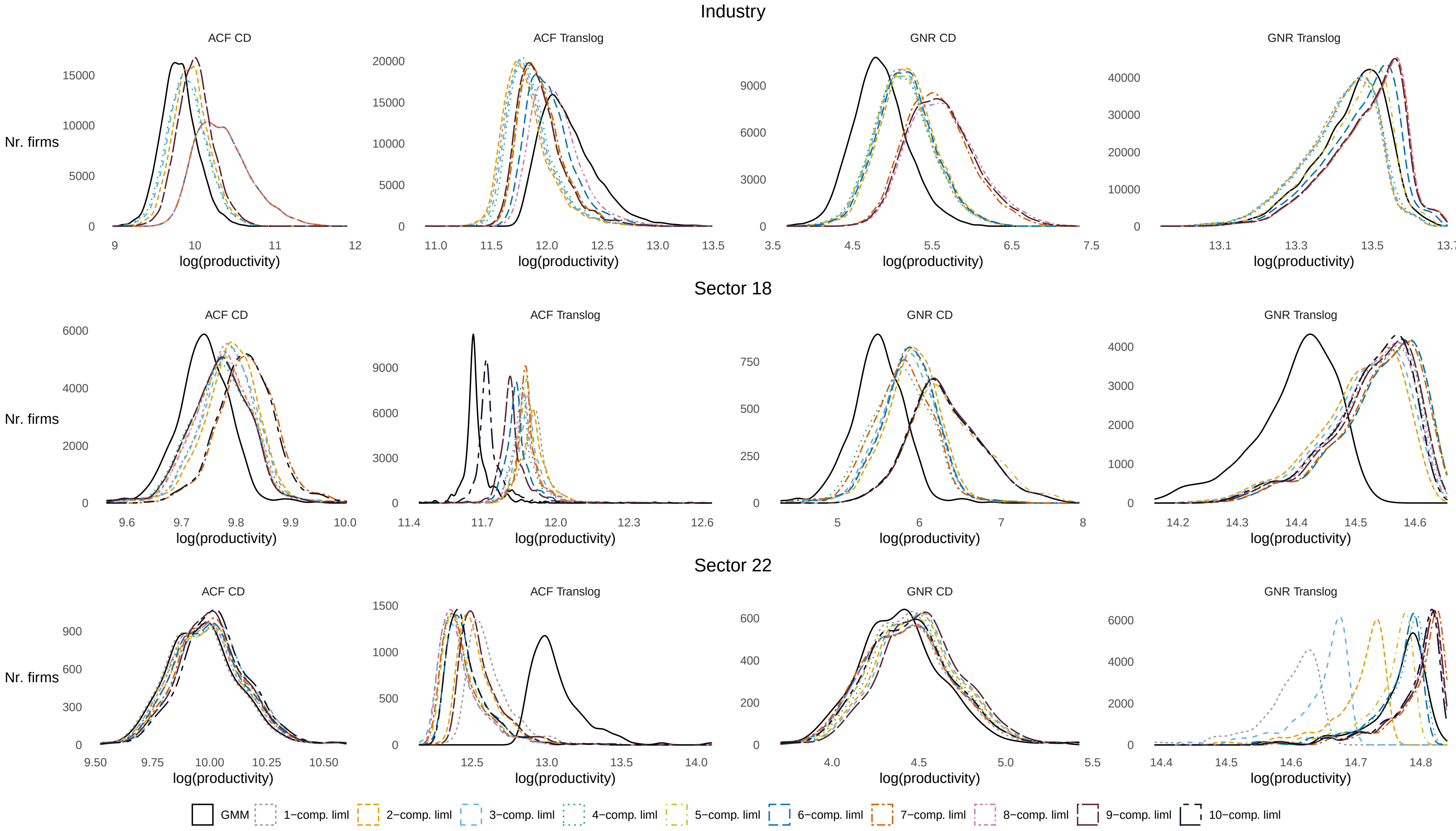

Figure 1: Density of 1-to 10-clustered productivity in 2013 obtained from Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with endogenous labor for the entire manufacturing sector and sectors 18 and 22 of the Belgian economy.

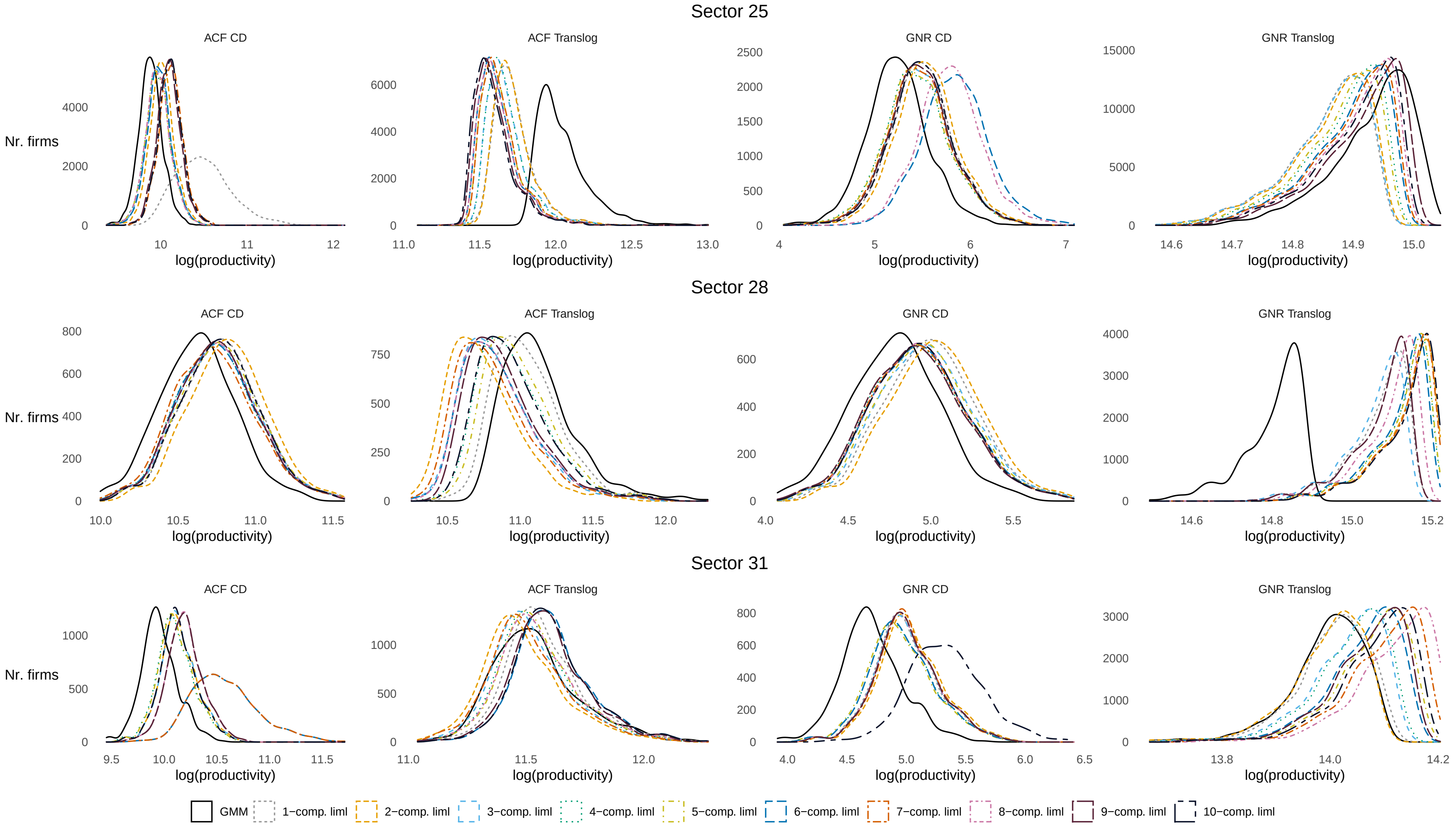

Figure 2: Density of 1-to 10-clustered productivity in 2013 obtained from Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with endogenous labor for sectors 25, 28, and 31 of the Belgian economy.

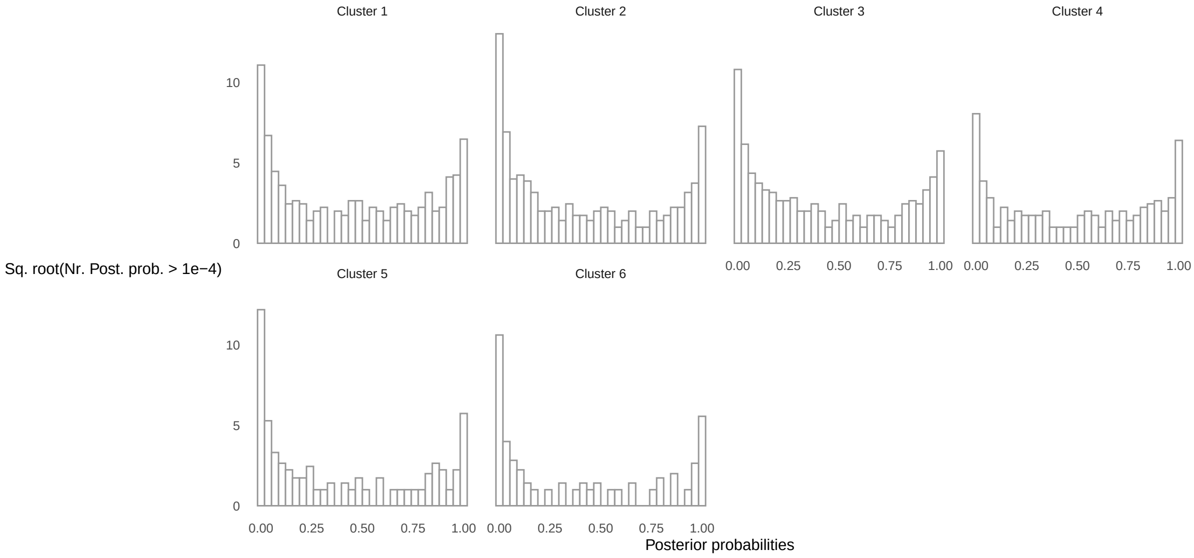

Figure 3: Histogram of posterior probabilities for a 6-cluster value-added Translog production function of NACE Rev. 22 estimated with LIML.

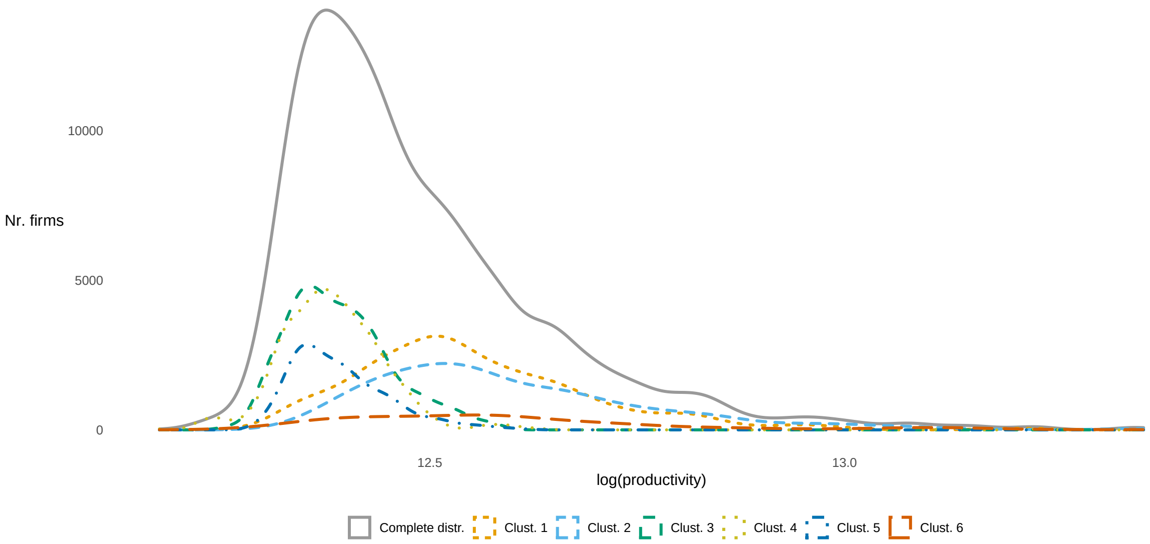

Figure 4: Complete and cluster-specific density of productivity in 2013 obtained from a 6-cluster value-added Translog production function of NACE Rev. 22 estimated with LIML.

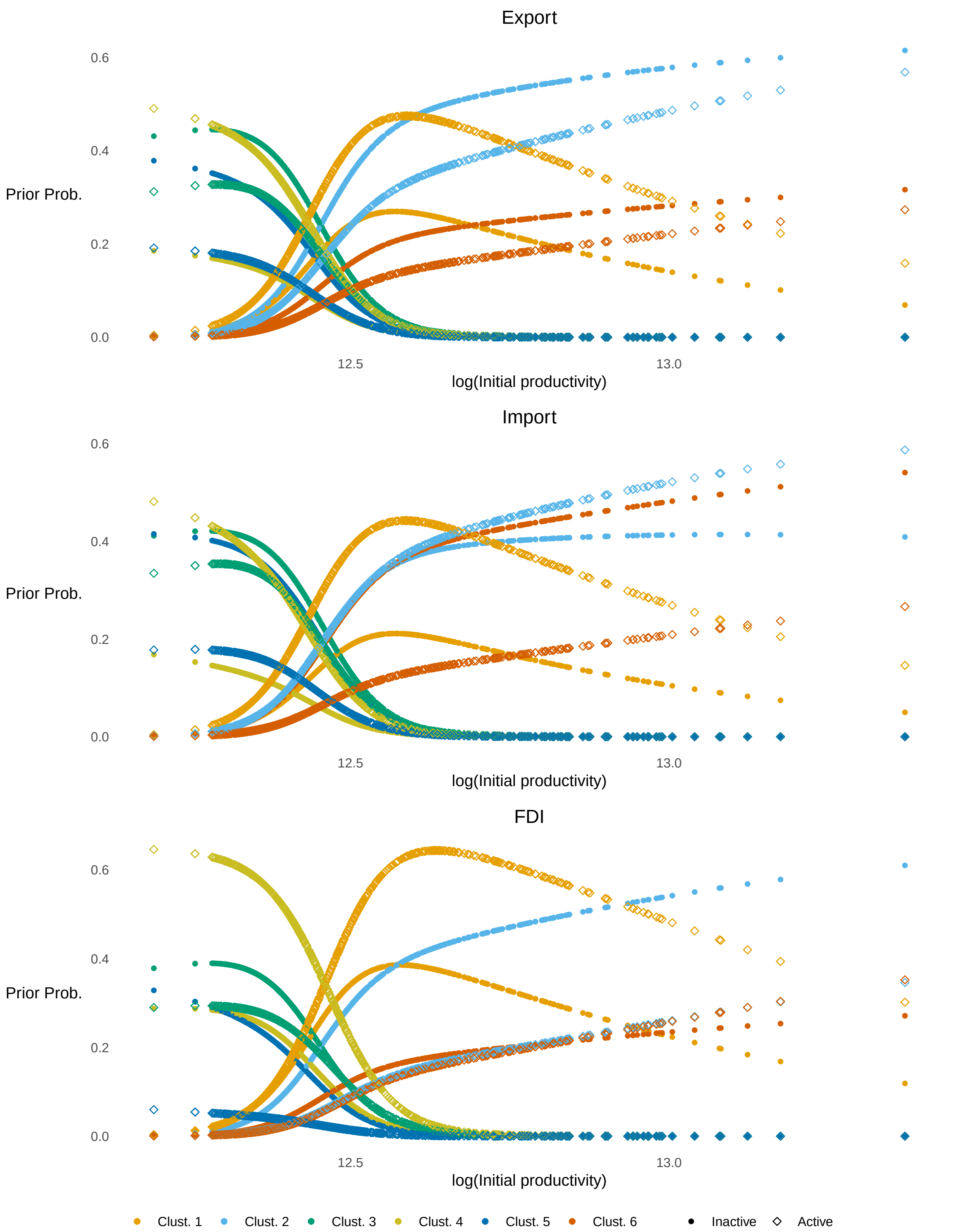

Figure 5: Cluster probability affiliation conditional as a function of an internationalization indicator (active as exporter, importer, or in FDI) and initial productivity.

Note: Cluster probability obtained from multinomial regression *after* production function estimation based on the following specification: $\frac{Pr(z_b^i|\ldots;\gamma^i)}{Pr(z_b^1|\ldots;\gamma^1)} = \gamma_0^i + \gamma_1^i \omega_{b0} + \gamma_5^i IntStatus_b, \quad \forall i = 2, \ldots, S$

Appendix B Productivity estimation

This section describes the production function estimation techniques relied upon in this paper. We summarize the proxy estimation techniques (Ackerberg et al., 2015) and first-order condition methods (Gandhi et al., 2020) before advancing our proposed Mixture Limited Information Maximum Likelihood estimator.

B.1 Proxy methods

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

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

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

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

Materials are assumed a flexible factor input, while capital and labor are considered non-flexible. As such, materials is chosen simultaneously 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 materials can function as a proxy for productivity:

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

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

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

The output from this first stage estimation allows us to define non-flexible output (log value-added) variation as:

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

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

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

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:

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

We parametrize equation 6 with production function coefficients $\boldsymbol{\beta}$ and specify the linear first-order Markov process, $g(\phi_{bt-1}, l_{bt-1}, k_{bt-1}, l_{bt}; \boldsymbol{\beta}, \boldsymbol{\alpha}) = \boldsymbol{W}_{bt-1}\boldsymbol{\alpha}^s$ with $\boldsymbol{W}_{bt-1} = [1, \omega_{bt-1}]$, and $\theta = \{\boldsymbol{\beta}, \boldsymbol{\alpha}\}$ such that

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

Specifying the instrumental variables as $\boldsymbol{Z}_{bt} = \left[\boldsymbol{Z}^{\boldsymbol{\beta}}_{bt}, \boldsymbol{W}_{bt-1}\right] = \left[k_{bt}, l_{bt(-1)}, \boldsymbol{W}_{bt-1}\right]$ and a weighting matrix $\left(\frac{\boldsymbol{Z}^T_{bt}\boldsymbol{Z}_{bt}}{BT}\right)^{-1}$, the optimization criterion is:

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

It can be observed the optimization problem is linear in the Markov process parameters $\boldsymbol{\alpha}$ and non-linear in the production function parameters $\boldsymbol{\beta}$. The corresponding First-Order Conditions (FOCs) are then:

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

   $\Leftrightarrow$

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

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

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

   $\Leftrightarrow$

   $$0 = \left(\frac{1}{BT}\sum_{b=1}^{B}\sum_{t=1}^{T}(\boldsymbol{Z}^{\boldsymbol{\beta}}_{bt})^T\nabla_{\boldsymbol{\beta}}\eta_{bt}\right)\left(\frac{(\boldsymbol{Z}^{\boldsymbol{\beta}}_{bt})\boldsymbol{Z}^{\boldsymbol{\beta}}_{bt}}{BT}\right)^{-1}\left(\frac{1}{BT}\sum_{b=1}^{B}\sum_{t=1}^{T}(\boldsymbol{Z}^{\boldsymbol{\beta}}_{bt})^T\eta_{bt}\right),$$

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

B.2 First-order condition methods

Gandhi et al. (2020) start from a gross output production function:

$$y_{bt} = f^{klm}(k_{bt}, l_{bt}, m_{bt}) + \omega_{bt} + \varepsilon_{bt}. \tag{10}$$

The estimator proposed by Gandhi et al. (2020), then, 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 elasticity of output with respect to materials and the ex-post Hicks-neutral productivity shock and possible measurement error ε_{bt}:

$$log\left(\frac{P^M M_{bt}}{P^Y Y_{bt}}\right) = log(\mathcal{E}) + log\left(\frac{\partial f^{klm}(k_{bt}, l_{bt}, m_{bt})}{\partial m_{bt}}\right) - \varepsilon_{bt} \tag{11}$$

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

$$\phi_{bt} = y_{bt} - \varepsilon_{bt} - \int \frac{\partial f^{klm}(k_{bt}, l_{bt}, m_{bt})}{\partial m_{bt}} dm_{bt} = -f^{kl}(k_{bt}, l_{bt}) + \omega_{bt}. \tag{12}$$

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

$$\phi_{bt} = -f^{kl}(k_{bt}, l_{bt}) + g\left(\phi_{bt-1} - f^{kl}(k_{bt-1}, l_{bt-1})\right) + \eta_{bt}. \tag{13}$$

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:

$$E\left[\eta_{bt} | k_{bt}, l_{bt(-1)}, \phi_{bt-1}\right] = 0. \tag{14}$$

We parametrize equation 13 with production function coefficients $\boldsymbol{\beta}$ and specify the linear first-order Markov process, $g(\phi_{bt-1}, l_{bt-1}, k_{bt-1}, l_{bt}; \boldsymbol{\beta}, \boldsymbol{\alpha}) = \boldsymbol{W}_{bt-1} \boldsymbol{\alpha}^s$ with $\boldsymbol{W}_{bt-1} = [1, \omega_{bt-1}]$, and $\theta = \{\boldsymbol{\beta}, \boldsymbol{\alpha}\}$ such that

$$\phi_{bt} = -f^{kl}(k_{bt}, l_{bt}; \boldsymbol{\beta}) + \boldsymbol{W}_{bt-1} \boldsymbol{\alpha} + \eta_{bt}. \tag{15}$$

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

B.3 Mixture (Limited Information) Maximum Likelihood

The methodology proposed in this paper builds on existing estimation methodologies for the first stage estimation (Ackerberg et al., 2015; Gandhi et al., 2020). These first-stage estimation procedures 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:

$$\mathcal{L}^c(\boldsymbol{\Theta}, \boldsymbol{z}) = \sum_{b=1}^{B}\sum_{s=1}^{S} z_b^s log\left(Pr(z_b^s|k_{b0}, l_{b0}, \omega_{b0}; \boldsymbol{\gamma}^s) p^o(\boldsymbol{\phi}_b, \boldsymbol{l}_b|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b, z_b^s; \boldsymbol{\theta}^s)\right). \tag{16}$$

We estimate the parameters of interest based on equation 16 relying on the Expectation-Maximization algorithm (McLachlan and Peel, 2000; Miljkovic and Grün, 2016). This algorithm consists of maximizing the complete log-likelihood in an iterative procedure. Assume parameter values in iteration j are represented as $(\boldsymbol{\Theta})^j = \{(\boldsymbol{\gamma}^1)^j, \ldots, (\boldsymbol{\gamma}^S)^j, (\boldsymbol{\theta}^1)^j, \ldots, (\boldsymbol{\theta}^S)^j\}$.

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

$$\hat{z}_b^s = Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; (\boldsymbol{\Theta})^j) = \frac{Pr(z_b^s|k_{b0}, l_{b0}, \omega_{b0}; (\boldsymbol{\gamma}^s)^j) p^o(\boldsymbol{\phi}_b|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b, z_b^s; (\boldsymbol{\theta}^s)^j)}{p^o(\boldsymbol{\phi}; (\boldsymbol{\Theta})^j)} \tag{17}$$

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

(i)

$$\max_{(\boldsymbol{\theta})^{j+1}} \Lambda(\boldsymbol{\theta}^{j+1}) = \max_{(\boldsymbol{\theta})^{j+1}} \sum_{b=1}^{B}\sum_{s=1}^{S} Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; (\boldsymbol{\Theta})^j) log\left(p^o(\boldsymbol{\phi}_b|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b, z_b^s; (\boldsymbol{\theta}^s)^{j+1})\right); \tag{18}$$

(ii)

$$\max_{(\boldsymbol{\gamma}^s)^{j+1}} \Lambda((\boldsymbol{\gamma}^s)^{j+1}) = \max_{(\boldsymbol{\gamma}^s)^{j+1}} \sum_{b=1}^{B}\sum_{s=1}^{S} Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; (\boldsymbol{\Theta})^j) log\left(Pr(z_b^s|k_{b0}, l_{b0}, \omega_{b0}; (\boldsymbol{\gamma}^s)^{j+1})\right). \tag{19}$$

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 rather than least-squares optimization (entropy = TRUE). The maximum likelihood estimation of the cluster-probability weighted observed log-likelihood ($\Lambda(\theta^{j+1})$), however, 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:

$$p^o(\phi_{bt}, l_{bt}|k_{bt}, l_{bt}, \phi_{bt-1}, l_{bt-1}, k_{bt-1}, z_b^s; \underbrace{\boldsymbol{\beta}, \boldsymbol{\alpha}^s, \boldsymbol{\delta}^s, \boldsymbol{\Sigma}^s}_{=\boldsymbol{\theta}^s}) = \frac{e^{-\frac{1}{2}\boldsymbol{\epsilon}^T(\boldsymbol{\Sigma}^s)^{-1}\boldsymbol{\epsilon}}}{\sqrt{(2\pi)^2|\boldsymbol{\Sigma}^s|}}, \tag{20}$$

where $\boldsymbol{\epsilon} = \begin{bmatrix} \phi_{bt} - f^{kl}\left(k_{bt}, l_{bt}; \boldsymbol{\beta}\right) - g(\phi_{bt-1}, l_{bt-1}, k_{bt-1}, l_{bt}; \boldsymbol{\beta}, \boldsymbol{\alpha^s}) \\ l_{bt} - \delta_0 - \delta_1 k_{bt} - \delta_2^s \phi_{bt-1} - \delta_3^s k_{bt-1} - \delta_4^s l_{bt-1} \end{bmatrix}$ 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 procudure, we rely on the observation that equation 20 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

$$p^o(\phi_{bt}|k_{bt}, l_{bt}, \phi_{bt-1}, l_{bt-1}, k_{bt-1}, z_b^s; \boldsymbol{\beta}, \boldsymbol{\alpha}^s, \sigma_\eta^s, \sigma_{\eta,\zeta}^s) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\left[\left(\sigma_\eta^s\right)^2 - \frac{(\sigma_{\eta,\zeta}^s)^2}{\left(\hat{\sigma}_\zeta^s\right)^2}\right]}} e^{-\frac{1}{2}\frac{\left(\eta_{bt}^s - \frac{\sigma_{\eta,\zeta}^s}{\left(\hat{\sigma}_\zeta^s\right)^2}\zeta_{bt}^s\right)^2}{(\sigma_\eta^s)^2 - \frac{\sigma_{\eta,\zeta}^s}{\left(\hat{\sigma}_\zeta^s\right)^2}}}, \tag{21}$$

and

$$p^o(l_{bt}|k_{bt}, \phi_{bt-1}, l_{bt-1}, k_{bt-1}, z_b^s; \boldsymbol{\delta}^s, \sigma_\zeta^s) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\left(\sigma_\zeta^s\right)^2}} e^{-\frac{1}{2}\left(\frac{\zeta_{bt}^s}{\sigma_\zeta^s}\right)^2}. \tag{22}$$

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

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

$$\Leftrightarrow$$

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

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

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

In a 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, $g(\phi_{bt-1},l_{bt-1},k_{bt-1},l_{bt};\boldsymbol{\beta},\boldsymbol{\alpha^s})=\boldsymbol{W}_{bt-1}\boldsymbol{\alpha}^s$ with $\boldsymbol{W}_{bt-1}=[1,\omega_{bt-1}]$. It can be observed that 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. $$\nabla_{\boldsymbol{\alpha}^s}\Lambda\left(\boldsymbol{\theta}\right)=0=-\frac{1}{\left(\sigma_\eta^s\right)^2-\frac{(\sigma_{\eta,\zeta}^s)^2}{\left(\sigma_\zeta^s\right)^2}}\sum_{b=1}^{B}Pr(z_b^s|\boldsymbol{k}_b,\boldsymbol{l}_b,\boldsymbol{\phi}_b;\boldsymbol{\Theta})\sum_{t=1}^{T}\left(\nabla_{\boldsymbol{\alpha}^s}\left(\eta_{bt}^s-\frac{\sigma_{\eta,\zeta}^s}{\left(\sigma_\zeta^s\right)^2}\zeta_{bt}^s\right)\right)^T\left(\eta_{bt}^s-\frac{\sigma_{\eta,\zeta}^s}{\left(\sigma_\zeta^s\right)^2}\zeta_{bt}^s\right)$$

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

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

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

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

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

$$\Leftrightarrow$$

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

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

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

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

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. This 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 computation time needed when increasing the number of clusters.

B.3.2 Labor as a predetermined input

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

$$p^o(\phi_{bt}|k_{bt}, l_{bt}, \phi_{bt-1}, l_{bt-1}, k_{bt-1}, z_b^s; \boldsymbol{\beta}, \boldsymbol{\alpha}^s, \sigma_\eta^s) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\left(\sigma_\eta^s\right)^2}} e^{-\frac{1}{2}\left(\frac{\eta_{bt}^s}{\sigma_\eta^s}\right)^2}. \tag{23}$$

The FOC are then:

1. $$\nabla_{\boldsymbol{\alpha}^s}\Lambda\left(\boldsymbol{\theta}\right) = 0 = -\frac{1}{\left(\sigma_\eta^s\right)^2}\sum_{b=1}^{B} Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \boldsymbol{\Theta}) \sum_{t=1}^{T} \boldsymbol{W}_{bt-1}^T\left(\omega_{bt} - \boldsymbol{W}_{bt-1}\boldsymbol{\alpha}^s\right)$$
$$\Leftrightarrow$$
$$0 = \sum_{b=1}^{B}\sum_{t=1}^{T} \boldsymbol{W}_{bt-1}^T Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \boldsymbol{\Theta})\left(\omega_{bt} - \boldsymbol{W}_{bt-1}\boldsymbol{\alpha}^s\right)$$
$$\boldsymbol{\alpha}^s = \left(\sum_{b=1}^{B}\sum_{t=1}^{T} \boldsymbol{W}_{bt-1}^T Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \boldsymbol{\Theta})\boldsymbol{W}_{bt-1}\right)^{-1} \sum_{b=1}^{B}\sum_{t=1}^{T} \boldsymbol{W}_{bt-1}^T Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \boldsymbol{\Theta})\omega_{bt};$$
2. $$\nabla_{\boldsymbol{\beta}}\Lambda\left(\boldsymbol{\theta}\right) = 0 = \sum_{s=1}^{S} -\frac{1}{\left(\sigma_\eta^s\right)^2}\sum_{b=1}^{B} Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \boldsymbol{\Theta}) \sum_{t=1}^{T} \nabla_{\boldsymbol{\beta}}(\eta_{bt}^s)\left(\omega_{bt} - \boldsymbol{\omega}_{bt-1}\boldsymbol{\alpha}^s\right)$$
$$\Leftrightarrow$$
$$0 = \sum_{s=1}^{S}\sum_{b=1}^{B}\sum_{t=1}^{T} \frac{1}{\left(\sigma_\eta^s\right)^2}\nabla_{\boldsymbol{\beta}}\left(\eta_{bt}^s\right) Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \boldsymbol{\Theta})\left(\omega_{bt} - \boldsymbol{\omega}_{bt-1}\boldsymbol{\alpha}^s\right),$$
where $\nabla_{\boldsymbol{\beta}}(\eta_{bt}^s) = -\nabla_{\boldsymbol{\beta}} f^{kl}\left(k_{bt}, l_{bt}; \boldsymbol{\beta}\right) + \alpha_2^s \nabla_{\boldsymbol{\beta}} f^{kl}\left(k_{bt-1}, l_{bt-1}; \boldsymbol{\beta}\right) - \boldsymbol{W}_{bt-1}\nabla_{\boldsymbol{\beta}}\alpha_2^s$;
3. $$\frac{\partial\Lambda\left(\boldsymbol{\theta}\right)}{\partial\left(\sigma_\eta^s\right)^2} = 0 = \sum_{b=1}^{B} Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \boldsymbol{\Theta})\left(-\frac{T}{2\left(\sigma_\eta^s\right)^2} + \frac{1}{2(\sigma_\eta^s)^4}\sum_{t=1}^{T}\left(\eta_{bt}^s\right)^2\right)$$
$$\Leftrightarrow$$
$$\left(\hat{\sigma}_\eta^s\right)^2 = \frac{\sum_{b=1}^{B}\sum_{t=1}^{T} Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \boldsymbol{\Theta})\left(\eta_{bt}^s\right)^2}{\sum_{b=1}^{B}\sum_{t=1}^{T} Pr(z_b^s|\boldsymbol{k}_b, \boldsymbol{l}_b, \boldsymbol{\phi}_b; \boldsymbol{\Theta})}.$$

Appendix C Monte Carlo

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

$$Y_{bt} = min\left\{K_{bt}^{\beta_k} L_{bt}^{\beta_l} e^{\omega_{bt}}, \beta_m M_{bt}\right\} e^{\epsilon_{bt}} \tag{24}$$

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, $\epsilon_{bt} \sim \mathcal{N}(0, 0.1)$. In contrast to Ackerberg et al. (2015), Log-productivity ω_{bt} follows an *Finite Mixture* AR(1)-process

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

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 shock to productivity follows 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 also Section 3.1). L_{bt} is chosen prior to period t without full knowledge of ω_{bt}. Strictly speaking, labor is chosen at time period $t-i$, with $i = 0.5$. We can think of decomposing the Finite Mixture AR(1)-process (eq. 25) into two subprocesses. First, ω_{bt-1} evolves to $\omega_{b,t-i}$, at which point in time the firm chooses labor input (as a function of $\omega_{b,t-i}$). Then, 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:

$$\begin{aligned} \omega_{b,t-i} &= \sum_{s=1}^{2} \left[\alpha_0^s + (\alpha_1^s)^{1-i} \omega_{b,t-1} + \eta_{bt}^{c,A}\right]^{z_b^s}; \\ \omega_{bt} &= \sum_{c=1}^{2} \left[\left(1 - (\alpha_1^s)^i\right) \alpha_0^s + (\alpha_1^s)^i \omega_{b,t-i} + \eta_{bt}^{c,B}\right]^{z_b^s}. \end{aligned} \tag{26}$$

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 equation 25 since $\left(1-(\alpha_1^s)^i\right)\alpha_0^s + \alpha_1^b\alpha_0^s = \alpha_0^s$ and $\left(\alpha_1^s\right)^{1-i}\left(\alpha_1^s\right)^i = \alpha_1^s$. Additionally, we follow Ackerberg et al. (2015) in imposing that $Var\left(\left(\alpha_1^s\right)^i \eta_{bt}^{s,A} + \eta_{bt}^{s,B}\right) = Var\left(\eta_{bt}^s\right)$ 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:

$$\ln\left(W_{bt}\right) = 0.3\ln\left(W_{bt-1}\right) + \eta_{bt}^{W}, \tag{27}$$

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 such 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):

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

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

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

$$K_{bt} = (1-\delta)K_{bt-1} + I_{bt-1}, \tag{28}$$

where $(1-\delta) = 0.8$. Investment is subject to convex adjustment costs given by

$$c_b\left(I_{bt}\right) = \frac{\phi_b}{2} I_{bt}^2, \tag{29}$$

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

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

$$
\begin{aligned}
I_{bt} =& \frac{\beta}{\phi_b} \sum_{\tau=0}^{\infty} (\beta(1-\delta))^{\tau} \left(\frac{\beta_k}{1-\beta_l} \right) \\
& \times \left[\beta_l^{\beta_l/(1-\beta_l)} - \beta_l^{1/(1-\beta_l)} \right] \\
& \times \exp \left\{ \left[\left(\frac{1}{1-\beta_l} \right) \alpha_0^c + \left(\frac{1}{1-\beta_l} \right) (\alpha_1^c)^{\tau+1} \omega_{bt} + \frac{-\beta_l}{1-\beta_l} \rho_W^{\tau+1} \ln (W_{bt}) \right. \right. \\
& + \frac{1}{2} \left(\frac{-\beta_l}{1-\beta_l} \right)^2 \sigma_{\eta^W}^2 \sum_{s=0}^{\tau} \rho_W^{2(\tau-s)} + \frac{1}{2} \left(\frac{1}{1-\beta_l} \right)^2 (\alpha_1^c)^{2b} \left((\alpha_1^c)^{2\tau} \sigma_{\eta^{c,A}}^2 \right. \\
& \left. \left. \left. + \sum_{s=1}^{\tau} \rho^{2(\tau-s)} \sigma_{\eta}^2 \right) + \left(\frac{1}{1-\beta_l} \right) \left(\frac{1}{2} \sigma_{\eta^{c,B}}^2 \right) \right] \right\}
\end{aligned}
\tag{30}
$$

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

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 sector 22. We demonstrate that the reported results are robust to estimation methodology and sector selection. We present the goodness-of-fit indicators (Online Appendix Figure 6), output elasticities and RTS (Online Appendix Figures 7 and 8), and measures of heterogeneity in the productivity distribution (Online Appendix Figures 9 and 10) for alternative estimation methodologies and all five sectors considered. It can be observed that the proposed method delivers reasonable estimates in all cases. Moreover, the stability of the output elasticities appears to be independent of the estimation methodology or selected sector. The Cobb-Douglas specifications are more volatile than the Translog specifications, but this volatility seems to originate from model misspecification or local maxima rather than underlying heterogeneity in the data. Only the value-added Translog specifications for the entire manufacturing industry and sector 28 demonstrate some signs of an omitted variable bias. However, the estimation results of the respective gross-output Translog specifications do not affirm this observation. Moreover, the stability of the proposed estimator to the addition of supplementary firm-level characteristics is also robust across industries. We supply in Online Appendix Table 1 the Log-likelihood, BIC, and ICLbic for different specifications of the cluster affiliation probabilities for all sectors. We observe that, independent of the sector, 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.

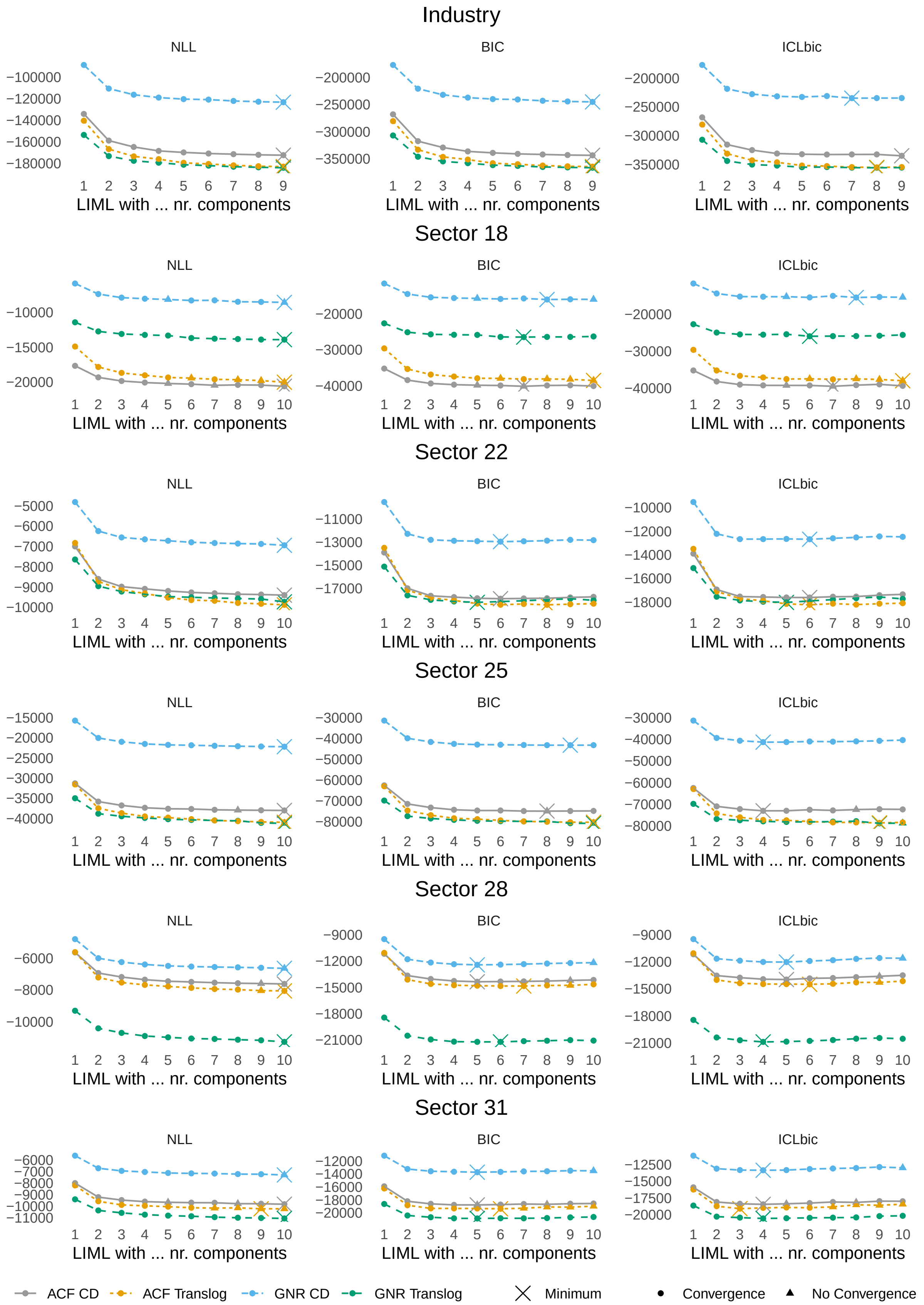

Figure 6: Evolution of Goodness-of-fit indicators of Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with endogenous labor in function of the number of clusters for the entire manufacturing sector and sectors 18, 22, 25, 28, and 31 of the Belgian 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 the non-convergence of the maximum likelihood estimation algorithm.

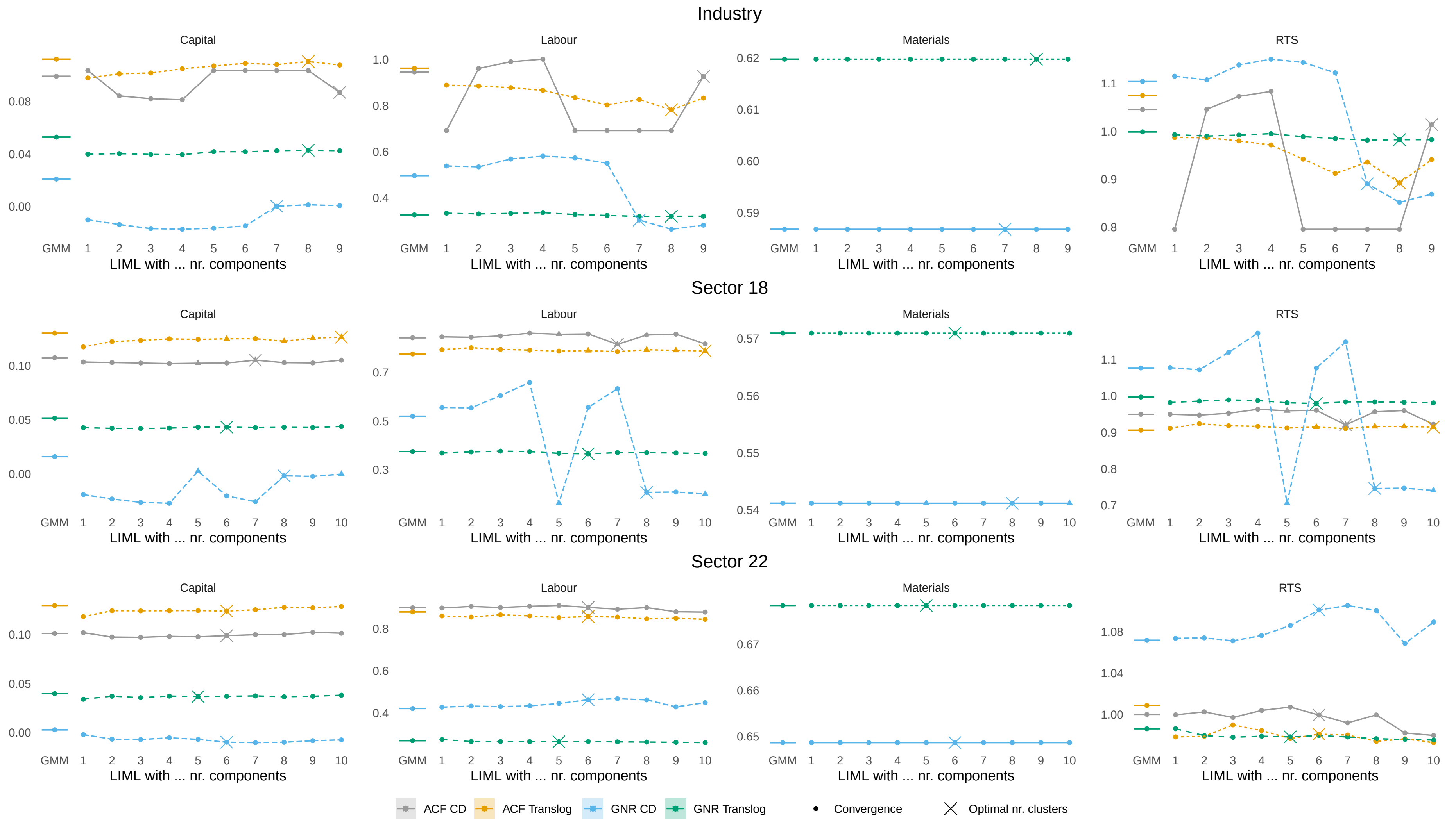

Figure 7: Evolution of output elasticities, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with endogenous labor, in function of the number of clusters for the entire manufacturing sector and sectors 18 and 22 of the Belgian economy.

Note: GMM and LIML refer to the reliance on the Generalised 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 the non-convergence of the maximum likelihood estimation algorithm.

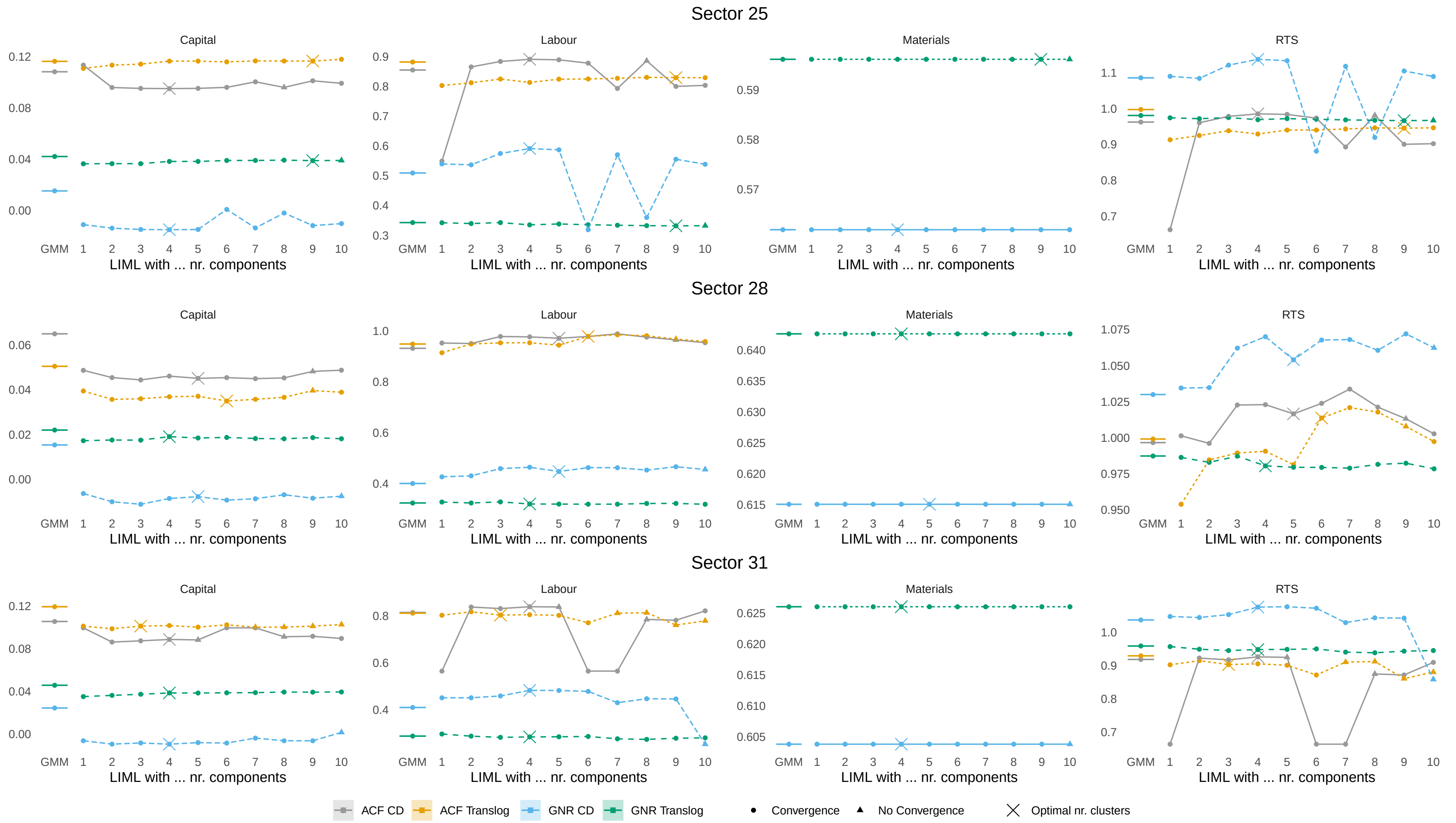

Figure 8: Evolution of output elasticities, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with endogenous labor, in function of the number of clusters for sectors 25, 28, and 31 of the Belgian economy.
Note: GMM and LIML refer to the reliance on the Generalised 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 the non-convergence of the maximum likelihood estimation algorithm.

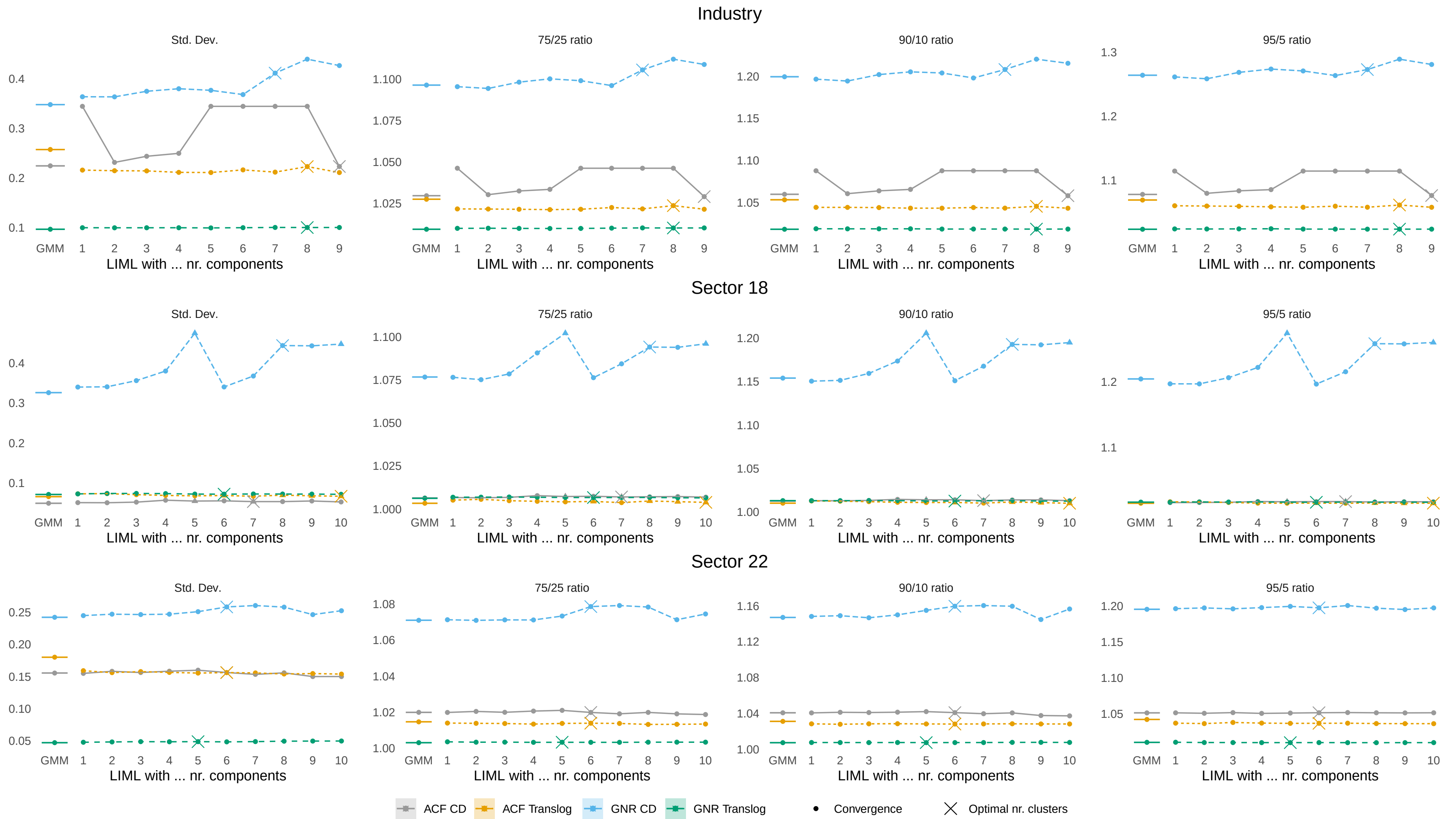

Figure 9: Evolution of the standard deviation, 75/25-, 90/10-, and 95/5-ratio, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with endogenous labor, in function of the number of clusters for the entire manufacturing sector and sectors 18 and 22 of the Belgian economy.

Note: GMM and LIML refer to the reliance on the Generalised 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 the non-convergence of the maximum likelihood estimation algorithm.

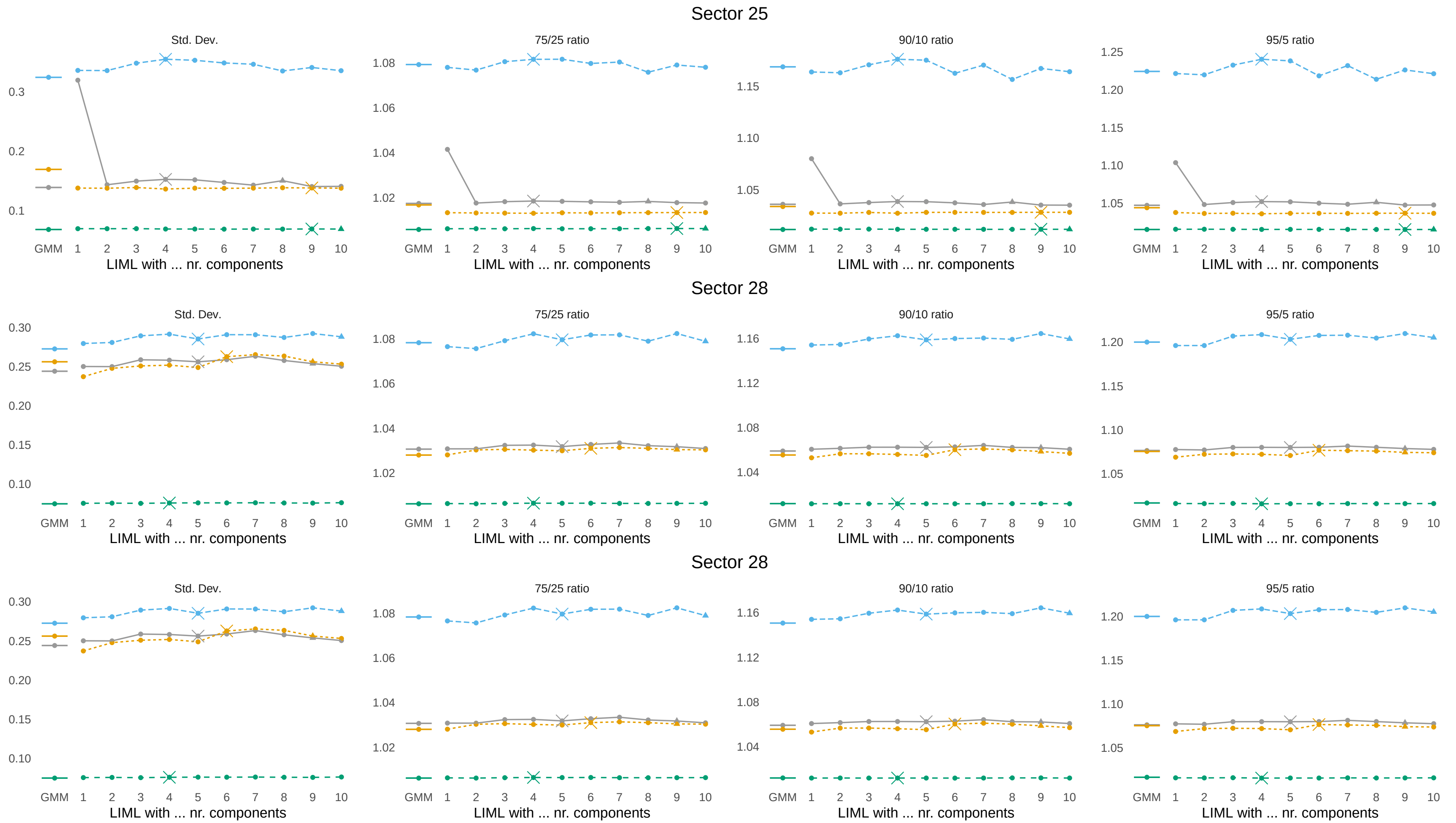

Figure 10: Evolution of the standard deviation, 75/25-, 90/10-, and 95/5-ratio, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with endogenous labor, in function of the number of clusters for sectors 25, 28, and 31 of the Belgian economy.

Note: GMM and LIML refer to the reliance on the Generalised 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 the non-convergence of the maximum likelihood estimation algorithm.

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

Specification	Log-likelihood	BIC	ICLbic
	Sector 18		
Base specification	20,155.61	-38,805.80	-38,326.01
Additional concomitants	20,117.91	-38,417.13	-37,924.09
Without initial capital and labor	19,958.99	-38,255.93	-37,724.51
	Sector 22		
Base specification	9,647.45	-18,433.93	-18,245.55
Additional concomitants	9,658.52	-18,292.07	-18,112.65
Without initial capital and labor	9,602.63	-18,262.30	-18,063.17
	Sector 25		
Base specification	40,894.37	-80,263.05	-78,635.87
Additional concomitants	40,963.76	-80,088.87	-78,495.14
Without initial capital and labor	40,779.89	-79,877.63	-78,165.77
	Sector 28		
Base specification	7,875.26	-14,864.30	-14,514.00
Additional concomitants	7,892.19	-14,729.36	-14,383.35
Without initial capital and labor	7,849.23	-14,727.85	-14,374.77
	Sector 31		
Base specification	9,895.65	-19,334.67	-19,117.96
Additional concomitants	9,903.90	-19,283.53	-19,070.44
Without initial capital and labor	9,828.85	-19,167.25	-18,934.84

Notes: a. The base specification refers to eq. 15, the augmented specification refers to eq. 35, and the specification without initial capital and labor refers to eq. 35 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 novel productivity estimation methodology and the robustness of the reported results for the Belgian manufacturing sector, we expand our estimation procedure to data on the Chilean manufacturing sector between 1979 and 1996 as relied on by (Gandhi et al., 2020) and provided by (Gandhi et al., 2020a). In line with the main results, the goodness of fit statistics displayed in Online Appendix Figure 13 provide evidence in favor of heterogeneity in productivity, the production function estimates shown in Online Appendix Figures 14 and 15 are close to those obtained with current state-of-the-art estimation methodologies, and the shape productivity distribution is not significantly affected when increasingly allowing for heterogeneity in productivity, as can be observed from the overall densities in Online Appendix Figures 11 and 12 and the summary statistics of the productivity distribution displayed in Online Appendix figures 16 and 17.

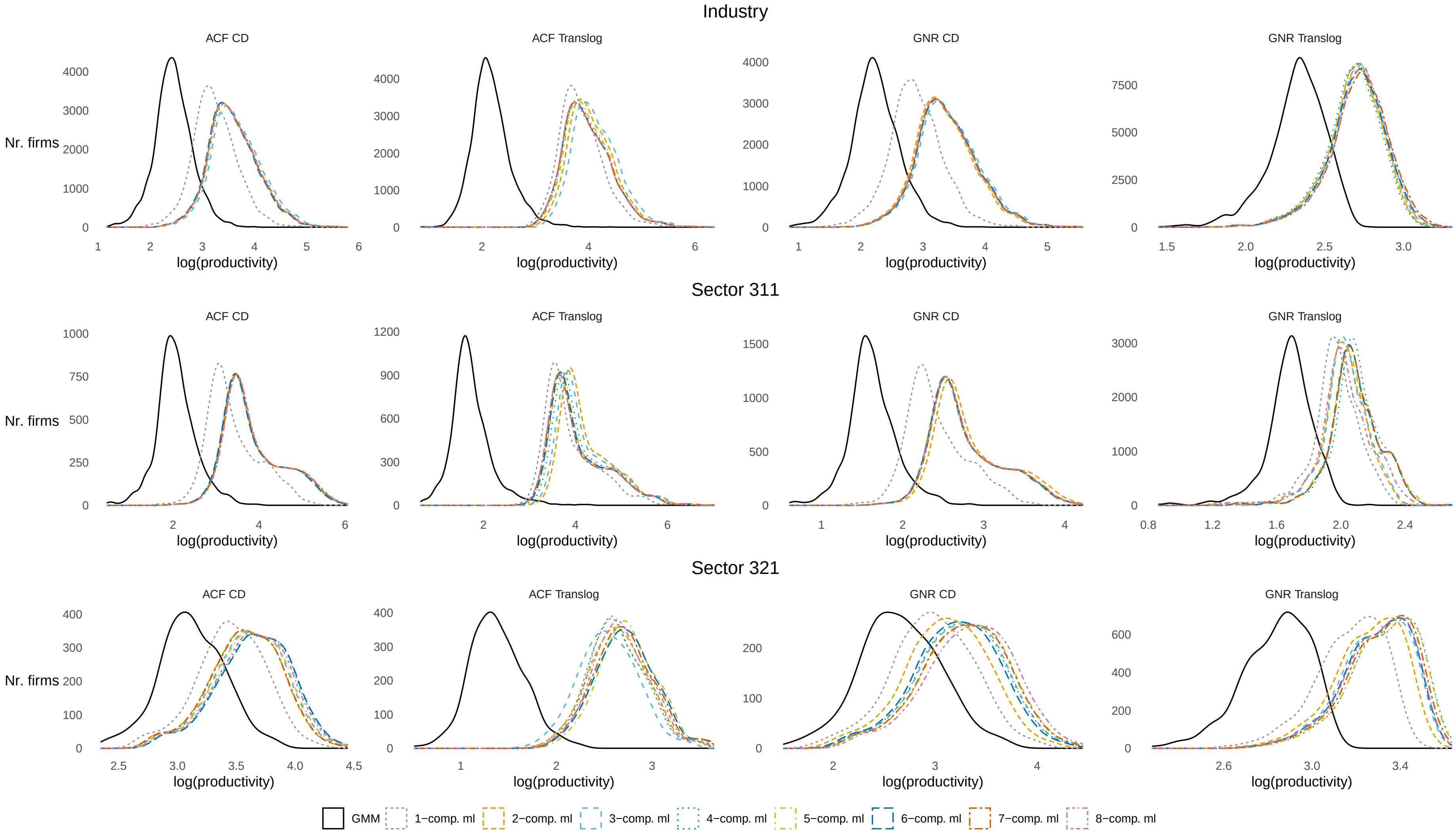

Figure 11: Density of 1-to 10-clustered productivity in 2013 obtained from Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with endogenous labor for the entire manufacturing sector and sectors 311, and 321 of the Chilean economy.

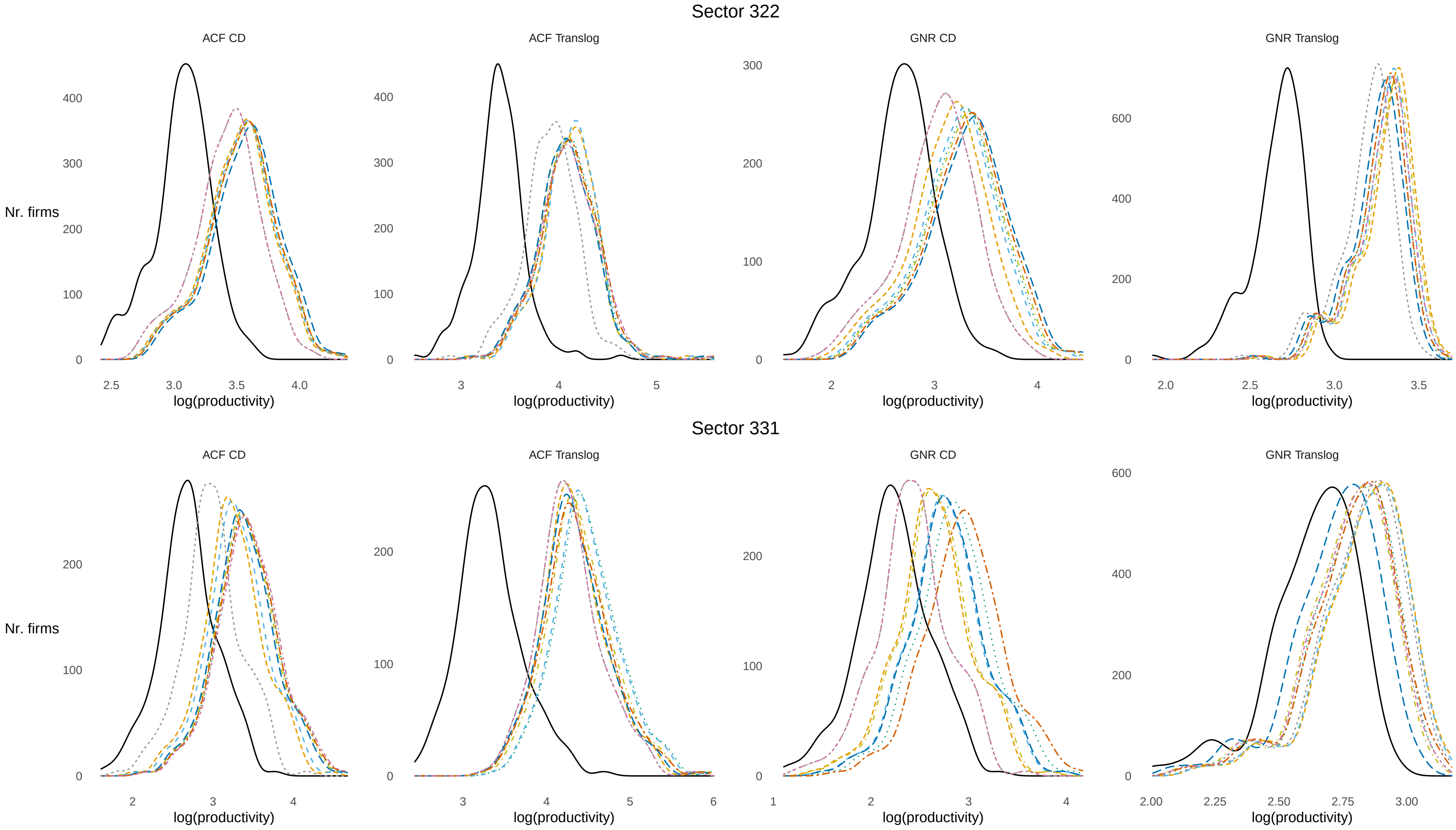

Figure 12: Density of 1-to 10-clustered productivity in 2013 obtained from Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with endogenous labor for sectors 322, 331, and 381 of the Chilean economy.

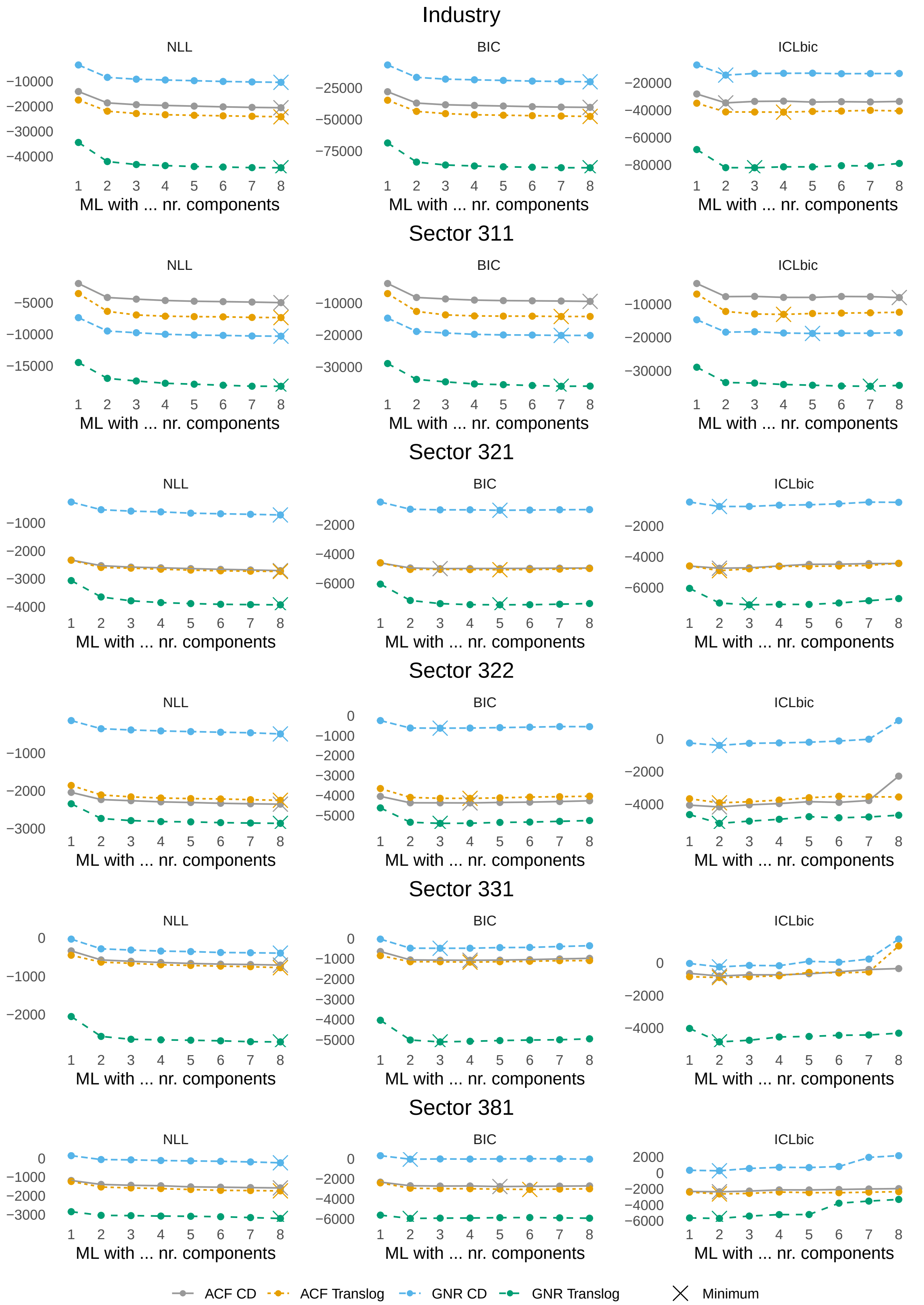

Figure 13: Evolution of Goodness-of-fit indicators of Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with exogenous labor, in function of the number of clusters for the complete Industry and sectors 311, 321, 322, 331, and 381 of the Chilean 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.

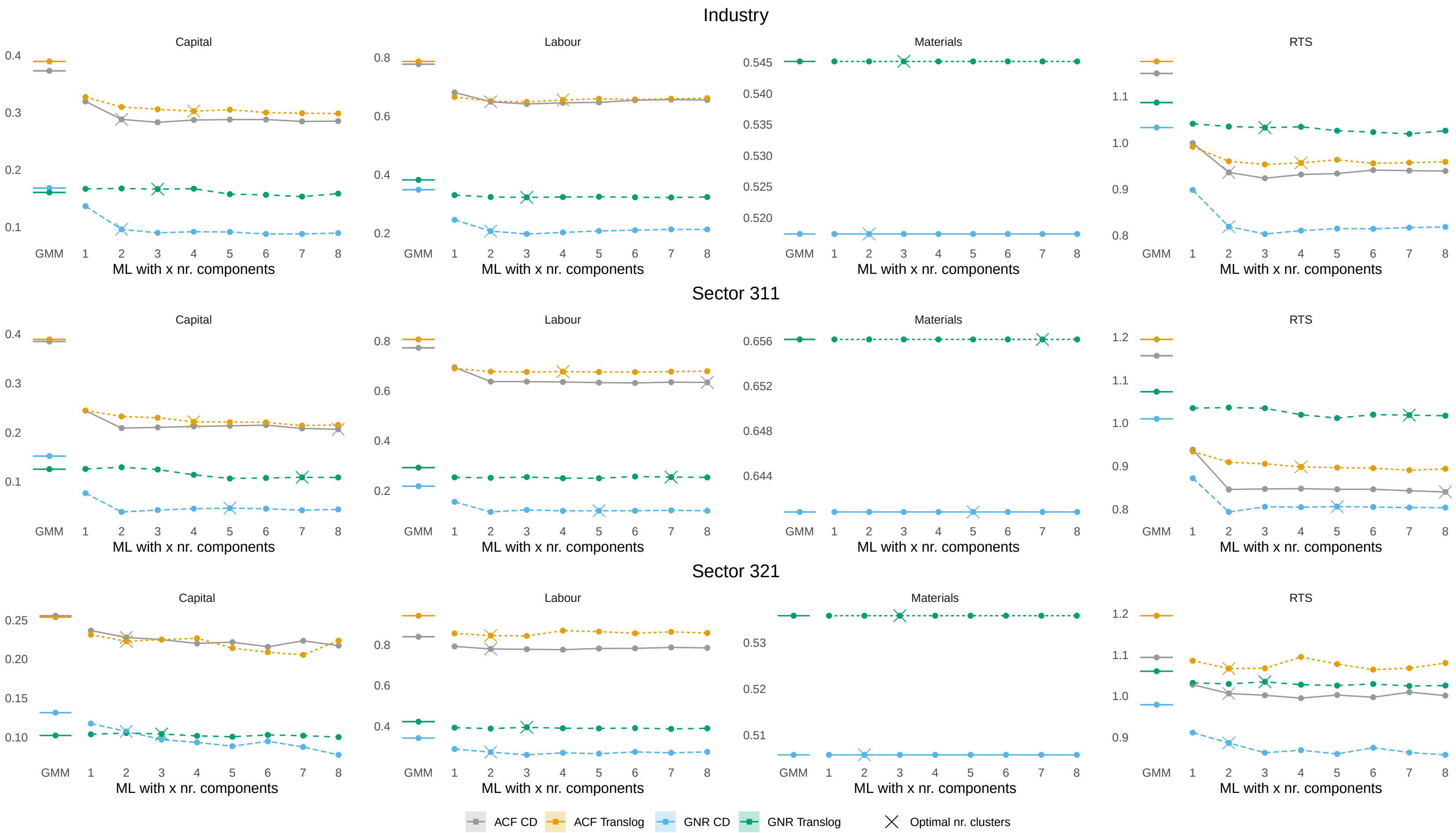

Figure 14: Evolution of output elasticities, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with exogenous labor, in function of the number of clusters for the complete Industry and sectors 311, 321 of the Chilean economy.

Note: GMM and ML refer to the reliance on the Generalised Method of Moments and Maximum Likelihood as estimation procedure. The times symbol indicates the optimal number of clusters defined by the Integrated Complete-data Likelihood Bayesian Information Criterion.

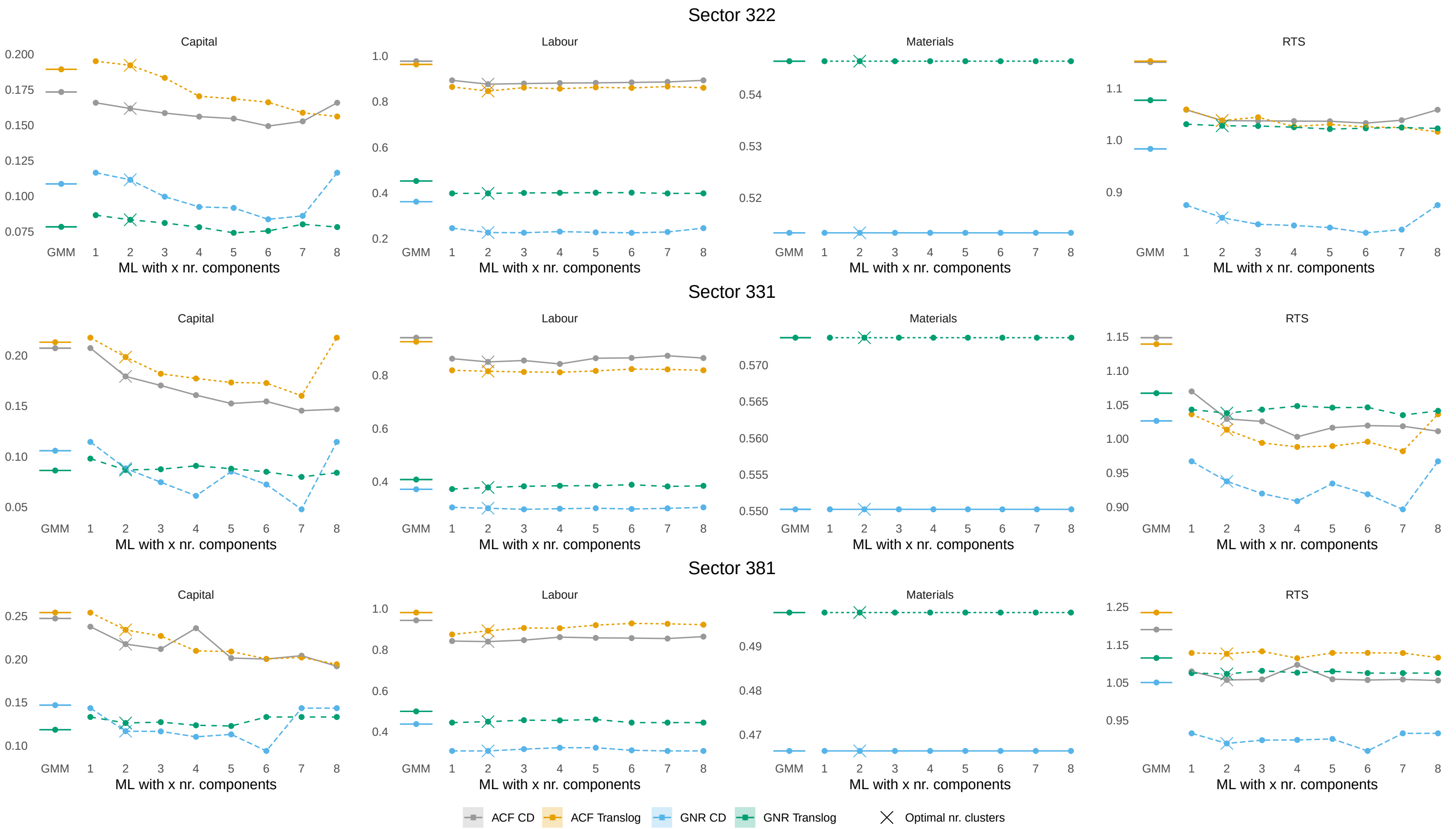

Figure 15: Evolution of output elasticities, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with exogenous labor, in function of the number of clusters for sectors 322, 331, and 381 of the Chilean economy.
Note: GMM and ML refer to the reliance on the Generalised Method of Moments and Maximum Likelihood as estimation procedure. The times symbol indicates the optimal number of clusters as defined by the Integrated Complete-data Likelihood Bayesian Information Criterion.

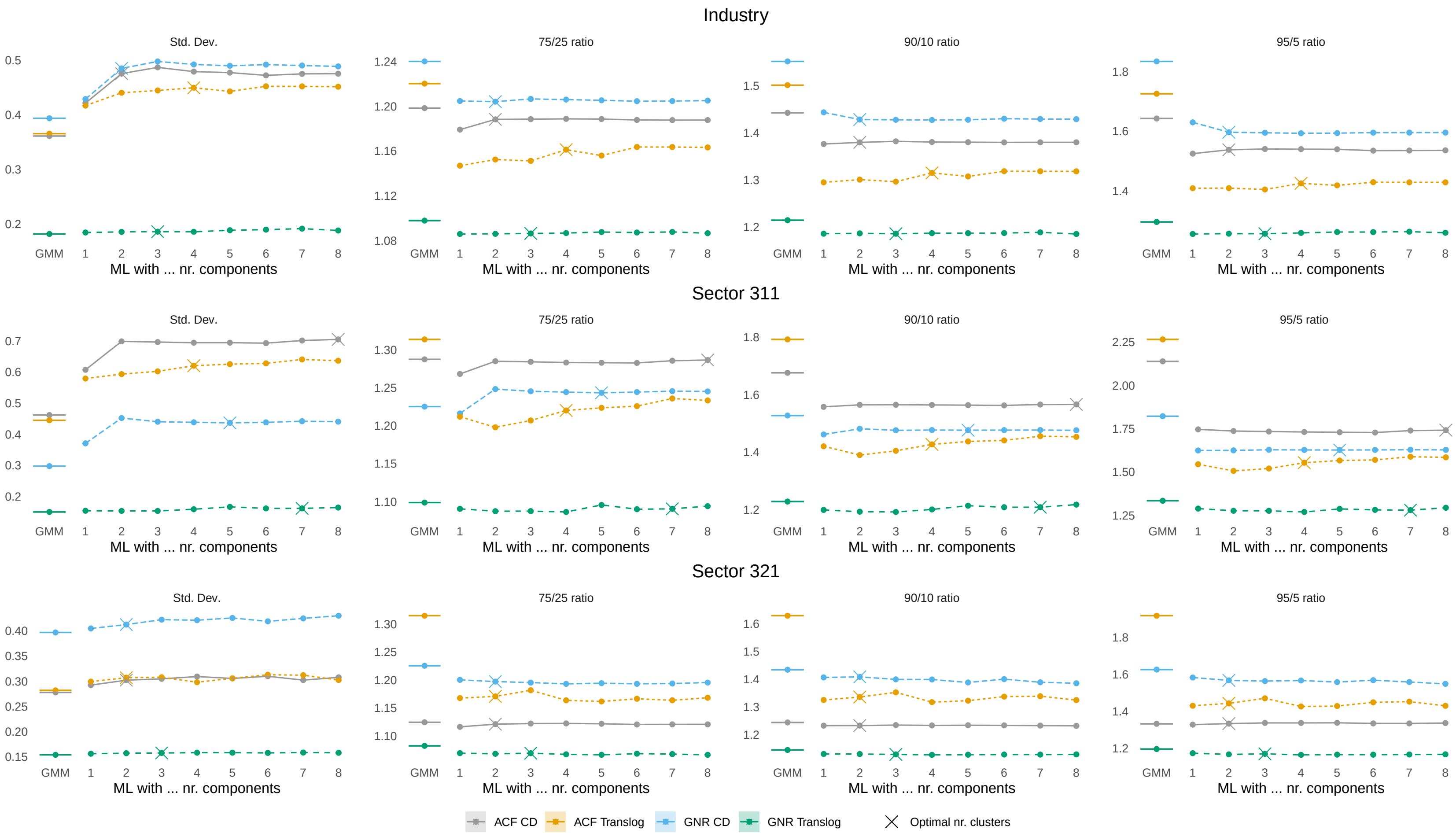

Figure 16: Evolution of the standard deviation, 75/25-, 90/10-, and 95/5-ratio, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with exogenous labor, in function of the number of clusters for the entire manufacturing sector and sectors 311 and 321 of the Chilean economy.

Note: GMM and ML refer to the reliance on the Generalised Method of Moments and Maximum Likelihood as estimation procedure. The times symbol indicates the optimal number of clusters defined by the Integrated Complete-data Likelihood Bayesian Information Criterion.

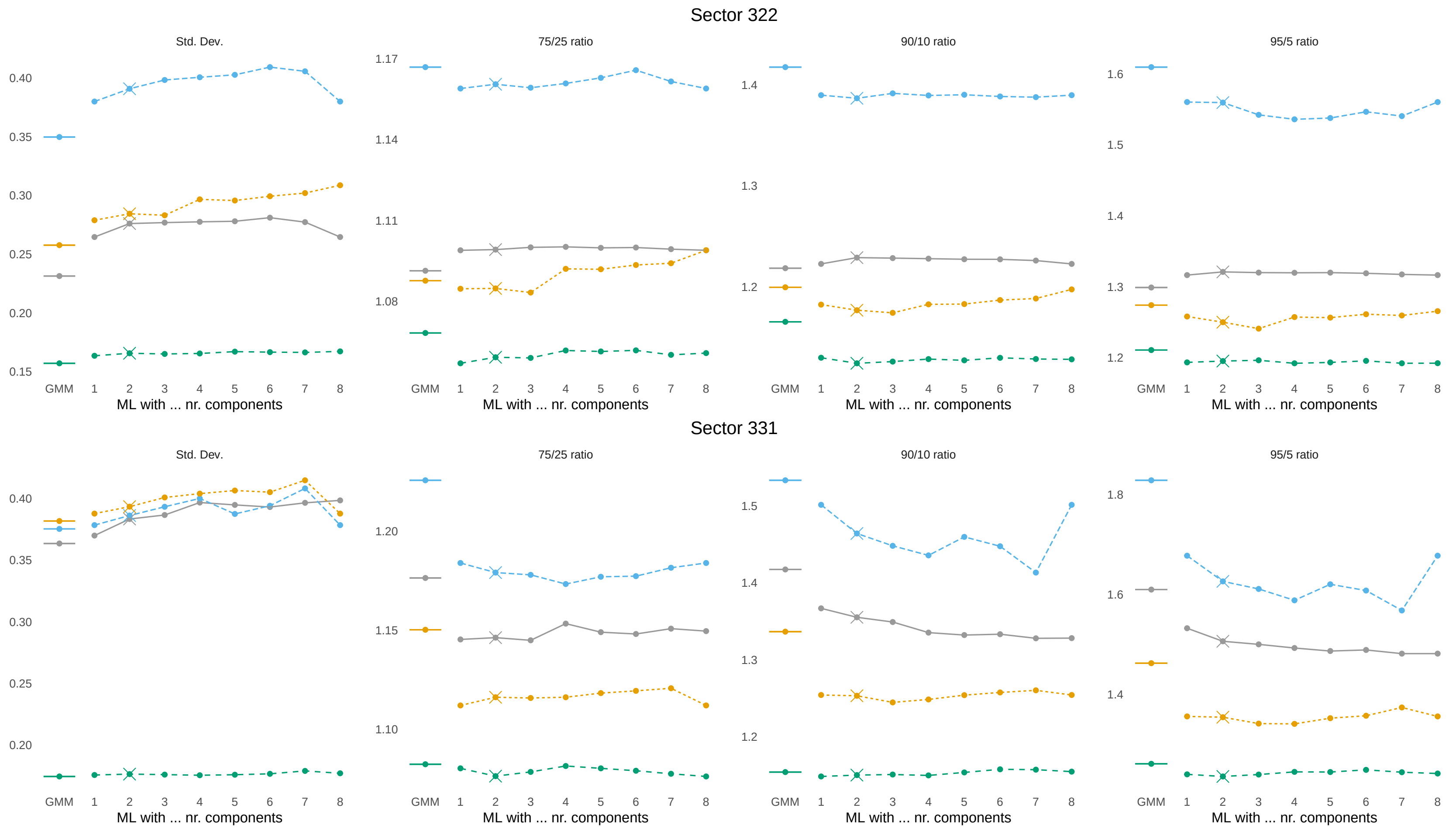

Figure 17: Evolution of the standard deviation, 75/25-, 90/10-, and 95/5-ratio, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with exogenous labor, in function of the number of clusters for sectors 321, 331, and 381 of the Chilean economy.

Note: GMM and ML refer to the reliance on the Generalised Method of Moments and Maximum Likelihood as estimation procedure. The times symbol indicates the optimal number of clusters defined by the Integrated Complete-data Likelihood Bayesian Information Criterion.

Appendix E Omitted Variable Bias

The limited evidence favoring an omitted variable bias in our main results is not in line with earlier findings reported in the literature (see Section 2). The roots of this deviation from the earlier literature can be traced back to methodological differences on two fronts.

First, the scalar unobservability assumption, stating that materials is a flexible factor input that is decided upon simultaneously at time t without affecting future profits ($m_{bt} = h(\omega_{bt}, k_{bt}, l_{bt})$), could be inadequate. Under the scalar unobservability assumption, the evolution of future productivity (and cluster affiliation) does not affect the choice of material inputs (Ackerberg, 2021). It can be argued, however, that cluster affiliation might affect the input demand of firms. A firm's export status, for instance, has been argued to lead to differences in optimal input demand across firms (De Loecker and Warzynski, 2012). If this export status is a determinant of cluster affiliation, 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. (2017) assuming the material input demand to be a function of 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 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 our second-stage production function estimation results' absence of an omitted variable bias. However, it is unclear why one would argue that the FOC for the perfectly flexible input would be cluster-dependent while FOCs of the non-flexible inputs are not. Cluster-dependent FOCs for all inputs implies a cluster-specific production function specification, which falls outside the scope of this paper. In the concluding section (Section 6), we discuss the possibilities the methodology proposed in this paper opens for future research, including cluster-dependent production function specifications.

Second, 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 ability of the deterministic cluster identification strategy to uncover an omitted variable bias in our Belgian firm-level data. 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 industry. 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) Cobb-Douglas and Translog specification. These production functions are estimated using a GMM estimation approach with either a simple linear Markov process specification:

$$\omega_{bt} = \alpha_0 + \alpha_1 \omega_{bt-1} + \eta_{bt}, \tag{31}$$

or a deterministic Markov specification:

$$\omega_{bt} = \alpha_0 + \alpha_1 \omega_{bt-1} + \alpha_2 D_b + \alpha_3 \left(\omega_{bt-1} \times D_b\right) + \eta_{bt}, \tag{32}$$

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, displayed in Figures 18 and 18, confirm the limited evidence in favor of an omitted variable bias 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', or 'all').

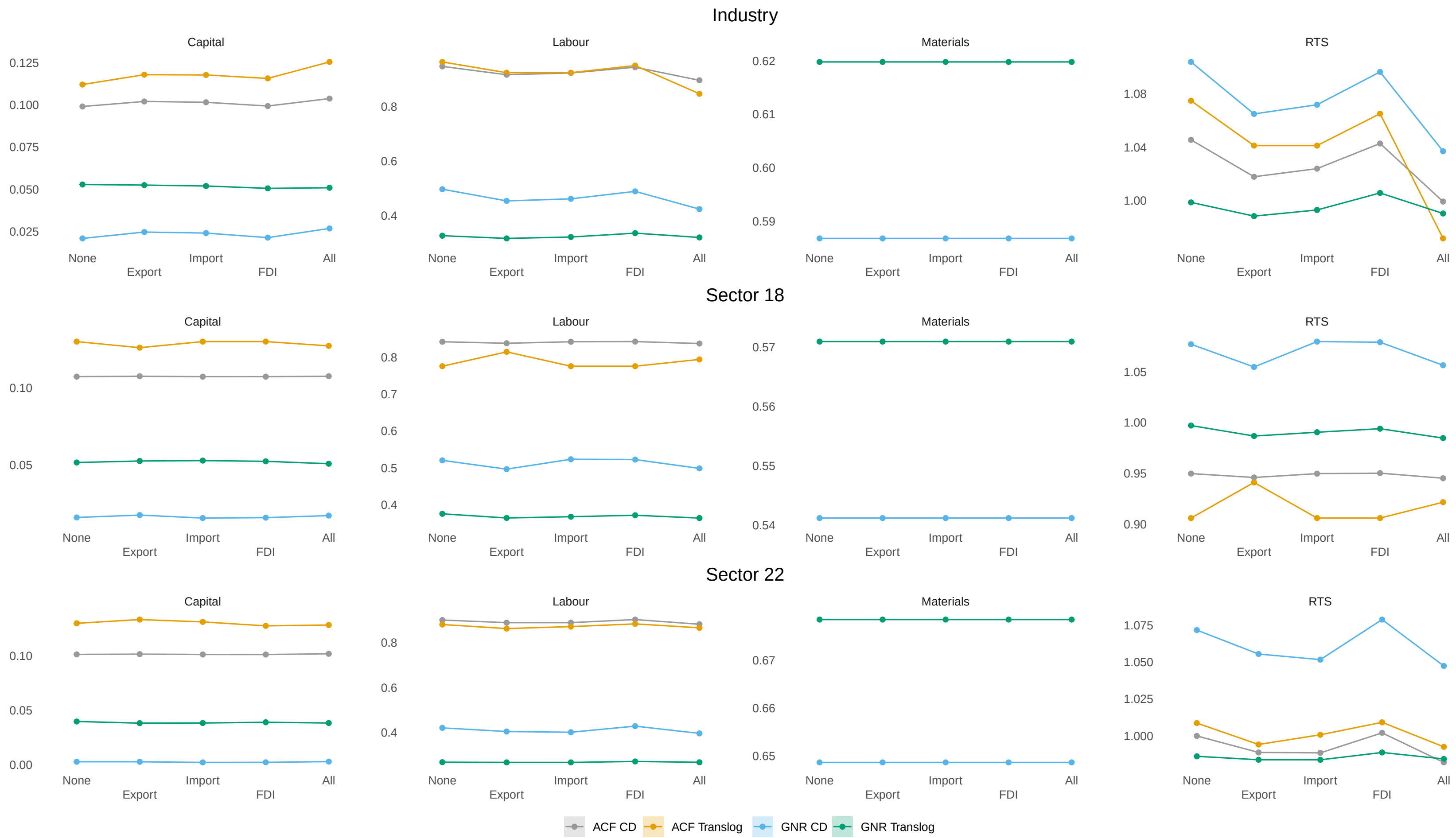

Figure 18: Evolution of output elasticities, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with endogenous labor, in function of the Markov specification for the complete Industry and sectors 18 and 22 of the Belgian economy.

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.

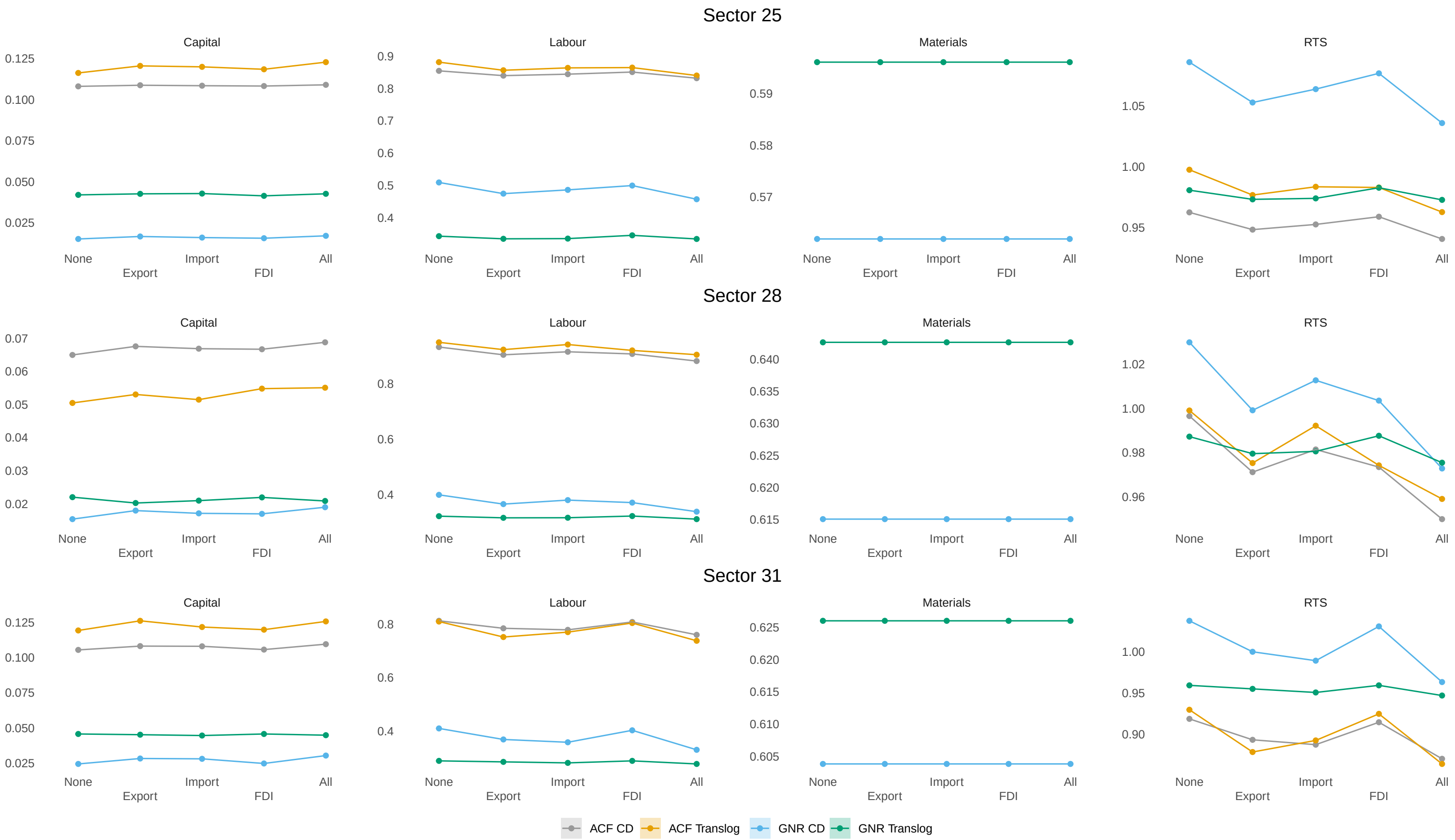

Figure 19: Evolution of output elasticities, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with endogenous labor, in function of the Markov specification for the sectors 25 and 28, and 31 of the Belgian economy.
Note: None, Export, 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.

Appendix References

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 (2017). Identification and estimation of production function with unobserved heterogeneity. *mimeo*.

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.