Identifying Latent Heterogeneity in Productivity

Ruben Dewitte, Catherine Fuss, and Angelos Theodorakopoulos*

January, 2024 [Click here for latest version]

Abstract

The determinants of firm-level productivity are often latent. The prevalence of these latent firm-level productivity determinants can lead to the mismeasurement of productivity differences between firm groups and their respective contribution to aggregate productivity growth. We propose a flexible extension of commonly employed production function estimation techniques using finite mixture models to control for latent factors. Monte Carlo evidence and estimates from Belgian and Chilean firm-level data demonstrate that controlling for such factors is crucial to obtaining accurate estimates of productivity differences between firm groups. Our approach delivers estimates of such productivity differences that are robust to the (un)availability of productivity determinants in the data.

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

JEL Codes: C13, C14, D24, L11

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

1 Introduction

Significant variation in firm-level productivity exists across different clusters of firms. The literature has pointed out several observable firm-level characteristics associated to these differences, including innovation (Aw et al., 2011; Doraszelski and Jaumandreu, 2013; Bilir and Morales, 2020), trade (Amiti and Konings, 2007; Kasahara and Rodrigue, 2008; De Loecker, 2013; Kasahara and Lapham, 2013; Merlevede and Theodorakopoulos, 2021), engagement in FDI (Javorcik, 2004; Blalock and Gertler, 2008), management practices (Bloom and Van Reenen, 2011; Caliendo et al., 2020; Rubens, 2020), technology (Harrigan et al., 2018), intangible transfers (Merlevede and Theodorakopoulos, 2020), human capital (Van Beveren and Vanormelingen, 2014; Konings and Vanormelingen, 2015) and industry linkages (Luttmer, 2007).

However, many characteristics affecting productivity remain latent and, thus, unobserved by researchers. Consequently, productivity differences between groups of firms can be misidentified when the underlying heterogeneity in the productivity growth process is left unexplained (De Loecker, 2013). This condition that all productivity determinants have to be accounted for imposes practically infeasible data requirements on a productivity estimation procedure. In particular, even if most heavy data requirements are met, some firm-level characteristics are expected to remain intrinsic and difficult to measure, e.g. managerial capacity and intangible capital (Haskel and Westlake, 2017).[1]

To illustrate the relevance of this observation, in Figure 1 we plot how the overall productivity density varies among firm clusters that are determined by firm-level characteristics available in the data, i.e. export and FDI status, for manufacturing firms both in Belgium in 2015 (left panel) and Chile in 1988 (right panel).[2] We find significant heterogeneity apparent in both the productivity average level and variance. Notably, Belgian firms engaged in FDI exhibit a distinct productivity growth path linked to a more dispersed and right-shifted productivity distribution compared to firms not involved in FDI. Although a similar pattern might be anticipated for Chilean firms, the data at hand lacks the necessary information to control for FDI status when estimating productivity. This logic can be, in turn, extended to all other latent productivity determinants not captured in the data.

This paper introduces a novel approach to estimating productivity that incorporates the influence of latent, time-invariant firm-level determinants on productivity growth

[1]From a technical point of view, considering a large set of relevant and possibly highly collinear explanatory variables raises the need for sufficient variation in the data to circumvent potential collinearity issues and obtain precise point estimates. This seems infeasible in practice, especially for small segments or sectors of the economy with a limited number of observations available by default.

[2]To estimate productivity, we follow De Loecker (2013) and directly incorporate firm-level characteristics into the productivity Markov process when estimating the production function, as detailed in Appendix F.

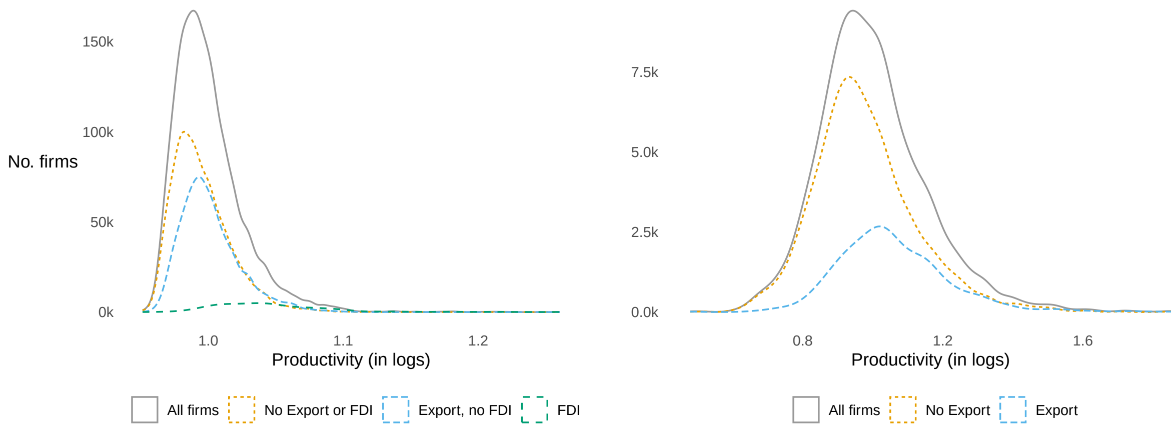

Figure 1: Productivity density by firm characteristics for the manufacturing sector in Belgium in 2015 (left panel) and Chile in 1988 (right panel)

Note: Productivity estimates are normalized to mean zero and obtained from a value-added Translog production function (Ackerberg et al., 2007) with the firm-level characteristics included in the Markov process. See sections 3 and 4 for an in-depth discussion of the productivity estimation methodology and underlying data.

heterogeneity. We build on the observation of Dewitte et al. (2022) that, in standard production function estimation methodologies, productivity for all firms is assumed to follow a homogeneous random growth process (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015; Gandhi et al., 2020). Therefore, if heterogeneity from firm-level characteristics is present but not controlled for, the level, dynamics, or drivers of productivity may be misidentified. To overcome these challenges, we propose an extension of current productivity estimation methodologies using finite mixture models (FMMs). A FMM is a probabilistic model that allows the productivity evolution to differ across clusters of firms, in line with the descriptive evidence presented in Figure 1, even when the underlying drivers of these differences remain unobserved by researchers. By employing FMMs, we capture the heterogeneity and account for the latent factors that influence productivity growth, thus enhancing the accuracy of productivity estimation.

The proposed methodology builds on the behavioral framework set out by Olley and Pakes (1996) and adapted by Levinsohn and Petrin (2003); Doraszelski and Jaumandreu (2013); Ackerberg et al. (2015), and Gandhi et al. (2020), among others. Specifically, as commonly modeled in the literature, a firm's production output is seen as a function of factor inputs and an additive Hicks-neutral productivity term. Instead of specifying productivity as the outcome of a growth process common to all firms, we allow it to evolve differently between clusters. We model the probability of cluster affiliation per firm without requiring additional information on the clusters' drivers besides—standard in this literature—information on firm-level output and input factors. We then extend the nonparametric identification arguments for the production function parameters (Gandhi et al., 2020; Ackerberg et al., 2022) to the cluster-specific parameters of the productivity growth process and the cluster affiliation probabilities by applying nonparametric identification results for finite mixtures of Markov processes (Kasahara and Shimotsu, 2009; Hu and Shum, 2012; Higgins and Jochmans, 2023a,b). Thus, by factoring latent heterogeneity into productivity, we can obtain "unbiased" estimates of productivity while reducing data requirements.[3]

We demonstrate the validity of the proposed methodology with a Monte Carlo analysis and, in turn, use detailed Belgian and Chilean firm-level data to showcase its empirical applicability. First, we extend the Monte Carlo experiment by Ackerberg et al. (2015) to account for latent heterogeneity in productivity. We highlight the benefits of the proposed method relative to current estimators when firm-level drivers of heterogeneity are unobserved in the data or affected by measurement error. Unlike aggregate productivity growth estimates, productivity differences between groups of firms, such as export premia, are biased when latent heterogeneity is not controlled for. Second, we show that the proposed estimator provides economically sensible estimates when brought to the data.[4] We apply our estimator to a set of Belgian manufacturing industries. We rely on data on firm-level revenue and input use from balance sheets and value-added tax (VAT) returns over the period 2008-2018, combined with a rich set of firm-level characteristics considered in the literature to be relevant for productivity growth, such as age, industry affiliation and participation in export, import, and foreign direct investment (FDI) activities. We find strong evidence of heterogeneity in the productivity growth process. We also observe multiple clusters of firms that differ in terms of this growth process, as reflected by differences in the level of productivity, the magnitude of unexpected shocks to productivity, and the persistence of these shocks over time.

The results suggest that cluster affiliation is positively associated with a firm's initial conditions, such as its initial productivity and factor input use. Any additional firm-level characteristics considered, such as age, export, import, or FDI status, can be associated with clusters but do not have strong explanatory power beyond the firm's initial conditions. This finding underlines the strength of the proposed estimation approach, whereby an unbiased identification of productivity does not necessarily require information on firm-level characteristics beyond output and input use. This aspect of the

[3]By unbiased estimates, we mean estimates that focus on resolving the identification of productivity in the presence of latent heterogeneity, without taking a stance on other forms of biases that may arise when estimating productivity. See Tybout (1992); Van Beveren (2012); De Loecker and Goldberg (2014) and De Loecker and Syverson (2021) for an overview. As such, the proposed estimation approach can be extended to incorporate existing solutions proposed in the literature to account for additional sources of bias that are likely to arise when estimating production functions.

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

proposed estimator is key in settings with limited data availability and in the study of complex economic environments where it is inherently difficult to single out key drivers of productivity growth.

To empirically demonstrate how the proposed estimator complements other commonly used methodologies that rely on information about firm-level characteristics, we focus on evaluating productivity differences between groups of firms.[5] To that end, we revisit a topic that has attracted the attention of multiple researchers and policymakers over the past years. Specifically, we examine the relative productivity advantage of exporting over non-exporting firms (the export premium), the evolution of this premium over time, and its contribution to aggregate productivity growth (see, for instance, Bernard and Bradford Jensen, 1999; Baldwin and Gu, 2003; Bernard et al., 2007; De Loecker, 2013; Garcia-Marin and Voigtländer, 2019; Gandhi et al., 2020). We demonstrate that export premia obtained from alternative productivity estimation methodologies can vary depending on the availability to the researcher of additional information on firm-level characteristics, such as age, import, or FDI status. In contrast, the proposed estimation approach delivers robust estimates of export premia and the contribution of exporting firms to aggregate productivity growth, regardless of the availability of additional information on firm-level characteristics in the dataset. A robustness check with Chilean firm-level data (Gandhi et al., 2020) reaffirms our baseline results based on Belgian data.

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

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

weak instruments.[6] The methodology proposed in this paper generalizes the productivity growth trend using FMMs, allowing for cluster-specific parameters (such as the constant) that account for unobserved heterogeneity while maintaining sufficient information to identify the production function parameters.

The advantages of FMMs have already been explored in the stochastic frontier literature (see, e.g., Beard et al., 1997; Orea and Kumbhakar, 2004; El-Gamal and Inanoglu, 2005; Greene, 2005) and in the structural production function literature to consider technology-specific production function specifications (Van Biesebroeck, 2003; Kasahara et al., 2017, 2023; Battisti et al., 2020). This paper shifts the focus to generalizing current productivity-estimation techniques, allowing for latent heterogeneity in the evolution of productivity. We generalize the estimation strategies for both value-added and gross-output production functions. To control for simultaneity problems, we build on the stochastic frontier literature (for an overview, see Amsler et al., 2016) to specify a (Limited Information) Maximum Likelihood ((LI)ML) model which does not require any additional assumptions regarding the first-order conditions of factor inputs beyond what is standard in the literature (Ackerberg et al., 2015; Gandhi et al., 2020). Moreover, we model the probability of belonging to a specific cluster. Such a mixture-of-experts specification (Gormley and Frühwirth-Schnatter, 2019) improves cluster identification and allows ex-post inference to be drawn for each cluster, for instance, by evaluating the correlation between identified clusters and firm-level characteristics. Compared to the stochastic frontier literature, the structural production function estimation techniques impose less stringent functional form restrictions (Sickles and Zelenyuk, 2019).

This paper is structured as follows. Section 2 introduces the model and establishes nonparametric identification for this model. We present and test our estimation framework by means of a Monte Carlo analysis in Section 3. We subsequently apply the methodology to firm-level data in Section 4, before discussing the robustness of the results in Section 5. We end with a summary of the main contributions and opportunities for future research in Section 6.

2 The model

In this section, we specify the assumptions underlying the production function estimation procedure following Gandhi et al. (2020) and Ackerberg et al. (2022). We then discuss the nonparametric identification of the production function provided by Gandhi et al. (2020) and Ackerberg et al. (2022) before extending their nonparametric identification

[6] To reduce such biases, one could further augment the estimation procedures borrowing from the 'system GMM' estimator developed by Blundell and Bond (1998) and outlined by Arellano and Bover (1995). For an application using the Gandhi et al. (2020) methodology, see Merlevede and Theodorakopoulos (2021).

strategy to our proposed generalization of the productivity process.

The data we observe consists of a (short) panel of firms $b = 1, \ldots, B$ over the period $t = 0, \ldots, T$, sampled from an underlying population. In each period t, firms have access to the information set $\mathcal{I}_{bt}$ when making their operating decisions. A generic firm's output, capital, labor, and intermediate inputs are denoted by $(Y_{bt}, K_{bt}, L_{bt}, M_{bt})$. This data allows the researcher to observe the joint distribution of $\{Y_{bt}, K_{bt}, L_{bt}, M_{bt}\}_{t=0}^{T}$. Furthermore, each firm belongs to a certain cluster.

Assumption 1 - Cluster affiliation: Each firm b belongs to a certain time-invariant cluster, indicated by $z_b \in \{1, \ldots, S\}$.[7] Cluster affiliation is dependent on the first-period information set, i.e., $Pr_z(z_b = s|\mathcal{I}_{b0})$ where $s = 1, \ldots, S$, and is known to the firm from that period onwards, $z_b \in \mathcal{I}_{bt}$ for $t > 0$. The total number of clusters S is exogenously determined.[8]

Specifying firm cluster membership as a function of the initial information set ensures that each firm does not have the same prior probability of following a particular cluster-specific productivity dynamic regardless of the firms' observable characteristics at entry. As will be demonstrated below, it also allows dealing with the initial conditions problem (Wooldridge, 2005; Frühwirth-Schnatter et al., 2012). Conditional cluster affiliation can, but does not have to, be economically motivated as the firm's optimal decision rule for cluster affiliation. This rule can follow from a net present value comparison between clusters and choosing the cluster that will result in the highest discounted profits, taking expectations and the costs of cluster affiliation into account.

2.1 The production function and cluster-specific productivity

The relation between output and inputs will be summarized by either a gross output production function (Gandhi et al., 2020), or a value-added production function (Ackerberg et al., 2015).

[7]Note that even though FMMs allow for time-invariant clusters, we limit the proposed approach to latent heterogeneity originating from time-invariant firm-level determinants in the productivity process to keep the resulting estimation method computationally fast and practical. This idea of fixed cluster membership over time is a stringent assumption, but not at odds with the Belgian firm-level data we use for our empirical application in the next section. Over ten years, 100% of the Belgian firms do not change their location, 91.9% do not change their industry affiliation, 81.8% do not change their export status, 72.6% do not change their import status, and 98.1% do not change their FDI status. Moreover, imminent changes to firm characteristics and resulting status can be expected to be related to their initial conditions and result in appropriately differentiated clusters.

[8]We use a general approach where the number of clusters is freely determined from the data.

Assumption 2a - Gross output production function: The relationship between output and inputs takes the form:

$$\begin{aligned} Y_{bt} &= f_{KLM}\left(K_{bt}, L_{bt}, M_{bt}\right) e^{\omega_{bt}+\varepsilon_{bt}} \qquad \Leftrightarrow \\ y_{bt} &= f_{klm}\left(k_{bt}, l_{bt}, m_{bt}\right) + \omega_{bt} + \varepsilon_{bt}, \end{aligned} \tag{1}$$

where lowercase variables indicate logarithmic values of uppercase variables. $f_{klm}(\cdot)$ represents the gross output production function explaining the variability in firm-level output, along with two additive terms.[9] The production function f_{klm} is differentiable at all $(k, l, m) \in \mathbb{R}^3_{++}$ and strictly concave in m.

Assumption 2b - Value added production function: The relationship between output and inputs takes the form (Gandhi et al., 2017; Ackerberg et al., 2015):

$$\begin{aligned} Y_{bt} - M_{bt} &= f_{KL}\left(K_{bt}, L_{bt}\right) e^{\omega_{bt}+\varepsilon_{bt}}, \qquad \Leftrightarrow \\ \frac{y_{bt}}{m_{bt}} &= f_{kl}\left(k_{bt}, l_{bt}\right) + \omega_{bt} + \varepsilon_{bt}, \end{aligned} \tag{2}$$

where $f_{kl}(\cdot)$ represents the value-added production function explaining the variability in firm-level output, along with two additive terms. The production function f_{kl} is differentiable at all $(k, l) \in \mathbb{R}^2_{++}$.

The two additive terms in equations (1) and (2) represent transitory and persistent shocks to measured production. The transitory component, ε_{bt}, represents an ex-post shock to production and possible classical measurement error that does not affect future output. The persistent component, ω_{bt}, represents Hicks-neutral total factor productivity (TFP) that is known to the firm before making its period t decisions. Concretely, ω_{bt} "might represent variables such as the managerial ability of a firm, expected downtime due to machine breakdown, expected defect rates in a manufacturing process, soil quality, or the expected rainfall at a particular farm's location", while ε_{bt} "might represent deviations from expected breakdown, defect, or rainfall amounts in a given year" (Ackerberg et al., 2015, p.2414). We formalize this interpretation as follows.

Assumption 3 - Shocks to production: $\varepsilon_{bt} \notin \mathcal{I}_{bt}$ is not known to the firm at the time of making its period t decisions and is independent of the within-period variation in information sets, so that its distribution can be written as $p_{\varepsilon}\left(\varepsilon_{bt} \mid \mathcal{I}_{bt}\right) = p_{\varepsilon}\left(\varepsilon_{bt}\right)$.[10]

[9] Throughout this paper, we differentiate between functional forms $f(\cdot)$ by indexing them with the input factors of interest x, i.e., $f_x(\cdot)$. Cluster-specific functional forms are indexed by the cluster-affiliation superscript s, i.e., $f^s_x(\cdot)$.

[10] Throughout this paper, we differentiate between discrete ($Pr(\cdot)$) and continuous ($p(\cdot)$) probability density functions (PDF) by indexing them with variable(s) of interest x, i.e., $Pr_x(x)$, $p_x(x)$. PDF's conditional on cluster affiliation are indexed by the cluster-affiliation indicator s, i.e., $p^s_x(x) = p_x(x|z_b = s)$.

$\omega_{bt} \in \mathcal{I}_{bt}$, on the other hand, is known to the firm at the time of making its period t decisions. Furthermore, ω_{jt} follows a *cluster-specific* first-order Markov process:

$$p_\omega(\omega_{bt}|\mathcal{I}_{bt-1}) = p_\omega\left(\omega_{bt}|\omega_{bt-1}, z_b\right). \tag{3}$$

This specification of the stochastic behavior of productivity generalizes commonly used specifications in the literature where either a single cluster is imposed, i.e. $S = 1$, or where a controlled Markov process is specified conditional on a vector of observed productivity determinants ($\boldsymbol{e}_{bt}$), i.e. $p_\omega(\omega_{bt}|\mathcal{I}_{bt-1}) = p_\omega(\omega_{bt}|\omega_{bt-1}, \boldsymbol{e}_{bt})$. In contrast, firm-level cluster affiliation, z_b, does not have to be observed by the econometrian a priori.

In combination with Assumption 1, Assumption 3 implies that we can decompose ω_{bt} into its cluster-specific conditional expectation and a cluster-specific innovation term[11]

$$\begin{aligned}\omega_{bt} &= \sum_{s=1}^{S} Pr_z(z_b = s|\mathcal{I}_{b0})\left(E\left[\omega_{bt}|\omega_{bt-1}, z_b = s\right] + \eta^s_{bt}\right)\\ &= \sum_{s=1}^{S} Pr_z(z_b = s|\mathcal{I}_{b0})\left(g^s(\omega_{bt-1}) + \eta^s_{bt}\right). \end{aligned} \tag{4}$$

where, by construction, $E_\eta\left[\sum_{s=1}^{S} Pr_z(z_b = s|\mathcal{I}_{b0})\eta^s_{bt}|\mathcal{I}_{bt-1}\right] = 0$ for $t \geq 1$.

The inputs, then, are decided upon according to the following rules.

Assumption 4 - Timing of input choices: Firms decide on period t capital in period $t-1$, such that capital is predetermined, i.e., $k_{bt} \in \mathcal{I}_{bt}$. Labor is determined either in period t, period $t-1$, or period $t-i$ with $0 < i < 1$. Both capital and labor input choices are dynamic such that $p_{k,l}(k_{bt}, l_{bt}|\omega_{bt-1}, k_{bt-1}, l_{bt-1})$ is dynamically complete. Materials are a fully flexible input, decided upon in period t based on the information set $\mathcal{I}_{bt}$, and are not affected by lagged values of materials.

2.2 The firm's problem

Assumption 5 - Market structure and profit maximization: Firms are price takers in the output and intermediate input markets, with P^M_t denoting the common intermediate-input price and P^Y_t denoting the common output price facing all firms in

[11] The proposed FMM specification contains a random coefficients specification and a multidimensional Hicks-neutral productivity specification as limiting cases. First, for S — the total number of clusters — going to B — the total number of firms — the FMM approach grows towards a random coefficients specification that allows for differences in the variance of the error term. Second, suppose productivity is a multidimensional sum of many independent random variables. In that case, (cluster-specific) productivity will be normally distributed as a result of the application of the Central Limit Theorem.

period t.[12] Firms maximize expected discounted profits.

Assumption 6 - First order condition of the flexible input: Under assumptions 2a, 4, and 5, the firm's profit-maximization problem with respect to intermediate inputs is

$$\max_{M_{bt}} P_t^Y E_\varepsilon \left[F_{klm} \left(K_{bt}, L_{bt}, M_{bt} \right) e^{\omega_{bt} + \varepsilon_{bt}} \mid \mathcal{I}_{bt} \right] - P_t^M M_{bt}, \tag{5}$$

as M_{bt} does not have any dynamic implications and thus affects only current-period profits. The first-order condition of the problem is

$$P_t^Y \frac{\partial}{\partial M_{bt}} F_{klm} \left(K_{bt}, L_{bt}, M_{bt} \right) e^{\omega_{bt}} E_\varepsilon \left[e^{\varepsilon_{bt}} \right] = P_t^M. \tag{6}$$

This equation can then be used to solve for the demand for intermediate inputs. The same logic applies in the value-added production function case under assumption 2b.

Assumption 7 - Flexible input demand: Under assumptions 4, 5, and respectively 2a (Gandhi et al., 2020) or 2b (Ackerberg et al., 2015), the firm's profit-maximization problem with respect to intermediate inputs motivates the definition of intermediate input demand as a function $\mathbb{M}_t (\cdot)$ of a single or, otherwise stated, scalar unobservable ω_{bt} which is strictly monotone in this unobservable ω_{bt}, and thus invertible:

$$m_{bt} = \mathbb{M}_t \left(k_{bt}, l_{bt}, \omega_{bt} \right) \quad \Leftrightarrow \quad \omega_{bt} = \mathbb{M}_t^{-1} \left(k_{bt}, l_{bt}, m_{bt} \right), \tag{7}$$

It should be noted that neither Assumption 6 nor Assumption 7 is affected by the generalization of the Markov process for productivity proposed in equations (3) and (4). This is because, under the stated assumptions, a flexible production factor will be unaffected by differences in the expectations of future productivity shocks between clusters of firms (Ackerberg, 2021).[13]

2.3 Nonparametric identification

Identification of the production functions specified in equations (1) and (2) is burdened by a simultaneity problem. Specifically, firm-level input choices depend on and thus correlate with the unobserved productivity term, i.e., $E \left[\omega_{bt} | k_{bt}, l_{bt}, m_{bt} \right] \neq 0$. This dependence

[12]For the sake of simplicity, we limit the behavioral framework to the case of perfect competition. However, the proposed identification procedure solely affects the assumption about the Markov process of productivity and can, therefore, naturally be extended to settings that allow for imperfectly competitive output and input markets, as in Klette and Griliches (1996); De Loecker (2011); Doraszelski and Jaumandreu (2013); De Loecker et al. (2016); Rubens (2021), and Blum et al. (2021).

[13]See Online Appendix F for a discussion of the adequacy of this assumption.

renders Ordinary Least Squares (OLS) or Nonlinear Least Squares (NLS) estimates of production function parameters inconsistent. Therefore, alternative identification strategies have been developed, usually consisting of two stages.

In the *first stage*, the ex-post production term (ε_{bt}) and the contribution of the flexible input factors are separated from output in the main estimating equation (1) or (2). Different methods exist to do so. For instance, Ackerberg et al. (2015) rely on the value-added production function (Assumption 2b) and the proportionality of the flexible production factor m_{bt} to value added (Assumptions 7). These assumptions allow the use of the flexible production factor, along with other variables, as a control for unobserved productivity to identify the ex-post shock to production and classical measurement error term ε_{bt}. On the other hand, Gandhi et al. (2020) build on the first-order conditions of the flexible production factor m_{bt} (Assumption 6) to jointly identify the ex-post shock to production term and the output elasticity of the flexible input from a gross output production function (Assumption 2a).[14] Both first-stage estimation procedures are consistent with the proposed generalization of the Markov process in this paper. As indicated below in subsection 3.2, both procedures rely on a flexible production factor unaffected by differences in the expectations of future productivity shocks between groups of firms (Ackerberg, 2021).

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

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

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

The *second stage* allows the identification of output elasticities of non-flexible inputs. It relies on the assumed Markov process of productivity (Assumption 4 and equation (4)) to replace the unobserved productivity term ω_{bt} from the above-specified equation (8) as a function of observables and production function parameters. The novel part of our methodology is that we generalize the productivity evolution process to explicitly depend on the fixed cluster affiliation of a firm through the cluster affiliation indicator z_b:

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

The question now is whether these proposed identification strategies allow for the pro-

[14]See Online Appendix C for a detailed description of the first stage of both methodologies.

duction function and, consequently, the productivity distribution to be identified nonparametrically based on the observed joint distribution of $\{y_{bt}, k_{bt}, l_{bt}, m_{bt}\}_{t=0}^{T}$, or $\{\phi_{bt}, k_{bt}, l_{bt}\}_{t=0}^{T}$ from the second stage onwards. If $S = 1$, under Assumptions 1–7 (respectively Assumption 2a and 2b) with the additional restriction on Assumption 4 that labor is predetermined ($l_{bt} \in \mathcal{I}_{bt}$), and an additional support condition on $(\phi_{bt}, k_{bt}, l_{bt})$, Gandhi et al. (2020) demonstrate that the gross output production function is nonparametrically identified up to an additive constant, while Ackerberg et al. (2022) demonstrate the nonparametric identification of the value-added production function up to an additive constant from the second stage onwards.[15]

To establish the nonparametric identification of the production function if $S > 1$, we write out the joint data distribution for the second stage as:[16]

$$p_{\phi,k,l,z}\left(\left\{\{\phi_{bt}, k_{bt}, l_{bt}\}_{t=0}^{T}, z_b\right\}_{b=1}^{B}\right) = \prod_{b=1}^{B}\sum_{s=1}^{S} Pr_z\left(z_b = s|\omega_{b0}, k_{b0}, l_{b0}\right) p_{\phi_{b0},k_{b0},l_{b0}}\left(\phi_{b0}, k_{b0}, l_{b0}\right) \times \prod_{t=1}^{T} p^{s}_{\phi,k,l}\left(\phi_{bt}, k_{bt}, l_{bt}|\phi_{bt-1}, k_{bt-1}, l_{bt-1}\right) \tag{10}$$

From this equality, it follows that the data admits a structure that allows to nonparametrically recover the conditional distribution of the latent clusters, the distribution of the initial conditions, and the cluster-specific data distribution up an arbitrary ordering of the clusters if $T \geq 6$ (Kasahara and Shimotsu, 2009), or $T \geq 4$ (Hu and Shum, 2012; Higgins and Jochmans, 2023a,b) based on the observed data distribution.[17] Upon identification of these elements, one can apply the identification arguments of Gandhi et al. (2020) for the gross output production structure or Ackerberg et al. (2022) for the value-added production structure to each cluster-specific data distribution separately to establish nonparametric identification of the production function and, consequently, the cluster-specific productivity distributions up to a cluster-specific additive constant.[18]

[15] Ackerberg et al. (2022) discuss the possibility of nonparametric identification in the case labor is dynamic but not predetermined, but argues that this requires stronger assumptions on the model structure than currently specified.

[16] See Online Appendix B for a detailed derivation. Note that we discuss identification for the second stage separately to clearly indicate that the first stage is cluster-independent and, therefore, nonparametrically identified independent of the identification of cluster affiliation.

[17] The identification arguments build on the multilinear restrictions that originate from the difference in each cluster's response pattern to variation in the covariate (Higgins and Jochmans, 2023a). The additional assumptions required for these arguments do not interfere with any of the previously discussed assumptions for nonparametric identification in the absence of firm clusters. Note that nonparametric identification of the conditional distribution of the latent clusters, the distribution of the initial conditions, and the cluster-specific data distribution is established for all possible labor input choices defined in Assumption 4.

[18] This process is similar to establishing nonparametric identification for a time-varying production function and Markov process by simply repeating the steps of the analysis separately for each time period (Gandhi et al., 2020, p.2994).

3 Estimation strategy

Below, we discuss the estimation strategy for the model structure specified above. We start by laying out the empirical specification and compare it to the dominant approach in the literature using Generalized Method of Moments (GMM). We discuss how we determine the optimal number of clusters and demonstrate the performance of the proposed estimation approach using a Monte Carlo exercise.

3.1 Empirical specification

The methodology proposed in this paper builds on existing two-stage estimation methods by directly applying 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 (see the model specification above and Ackerberg, 2021).

The second-stage estimation equation (9) simplifies to a standard equation that can be estimated with GMM if $S = 1$ (Ackerberg et al., 2015; Gandhi et al., 2020). When $S > 1$, equation (9) contains a cluster affiliation indicator that is latent to the researcher. While our model structure is nonparametrically identified, we resort to parametric restrictions in our estimation specification to keep the estimation procedure computationally feasible and practical. We demonstrate in the next subsection, 3.2, that these parametric restrictions are sufficiently general to approximate a nonparametric approach.

Assumption 8 - Productivity parametrization: We assume productivity follows a Gaussian mixture.

The assumption that productivity follows a Gaussian mixture is shared by Ackerberg et al. (2015); Kasahara et al. (2023), and is in line with Dewitte et al. (2022) who demonstrate that firm-size distribution is best represented by a finite mixture of log-normals.[19]

[19] Aside from empirical evidence, two arguments favor the (log-)normal specification of productivity. First, from the perspective of overall fit, a mixture of normal distributions with sufficient components is shown to be able to approach all distributions (McLachlan and Peel, 2000). This argument implies, however, that the number of mixtures does not necessarily coincide with the number of clusters in the data. Second, from a generative perspective for individual components, the normal distribution is the realization of applying the central limit theorem, whereby firm productivity is approximately normally distributed if it is the sum of many independent random variables. This corresponds to the multi-dimensional definition of productivity, for example, 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).

Assumption 9 - Cluster affiliation parametrization: We model the conditional probability of belonging to a specific cluster as a multinomial logit for z_b:

$$Pr_z(z_b = s|\omega_{b0}, k_{b0}, l_{b0}; \boldsymbol{\gamma}^1, \ldots, \boldsymbol{\gamma}^s) = \frac{e^{\gamma_0^s + \gamma_k^s k_{b0} + \gamma_l^s l_{b0} + \gamma_\omega^s \omega_{b0}}}{\sum_{i=1}^{S} e^{\gamma_0^i + \gamma_k^i k_{b0} + \gamma_l^i l_{b0} + \gamma_\omega^i \omega_{b0}}}, \qquad \forall s = 1, \ldots, S. \quad (11)$$

where $\boldsymbol{\gamma^s} \equiv \{\gamma_0^s, \gamma_k^s, \gamma_l^s, \gamma_\omega^s\}$ is a cluster-specific vector gathering the parameters of the multinomial logit model.

Lastly, we restrict Assumption 4 further regarding the labor input choice.

Restriction A of Assumption 4 - Timing of labor input choice: Labor is decided upon at time $t-1$ and, therefore, predetermined $l_{bt} \in \mathcal{I}_{bt}$.

Under the additional parametric assumptions imposed by Assumptions 8-9 and Restriction A of Assumption 4 on the timing of labor input choice, we specify the Maximum Likelihood (ML) equation ($\mathcal{L}_A(\cdot)$), i.e., the partial log-likelihood to be maximized based on equation (10), as follows:

$$\mathcal{L}_A(\boldsymbol{\Theta}) = \sum_{b=1}^{B} log \left(\sum_{s=1}^{S} Pr_z(z_b = s|\omega_{b0}, k_{b0}, l_{b0}, ; \boldsymbol{\gamma}) \prod_{t=1}^{T} p_\phi^s(\phi_{bt}|l_{bt}, k_{bt}, \phi_{bt-1}, k_{bt-1}, l_{bt-1}; \boldsymbol{\theta}^s) \right), \quad (12)$$

where $\boldsymbol{\theta}^s$ is a cluster-specific vector that captures the parameters of the productivity process $g_s(\cdot)$ and the production function $h_{kl}(\cdot)$, and $\boldsymbol{\Theta} \equiv \{\boldsymbol{\gamma}^1, \ldots, \boldsymbol{\gamma}^S, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^S\}$. By only maximizing over the partial likelihood, we avoid specifying the marginal model for the initial dependent variables or specifying a model for k_{bt} and l_{bt}.[20] Even though this implies losing information, we simplify the analysis and gain robustness because we do not have to specify these additional components explicitly. This is an advantage that our approach shares with the commonly employed GMM approach (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015; Gandhi et al., 2020), as inputs may be chosen as part of a complex dynamic optimization process (Ackerberg et al., 2015).

Labor, however, may be a dynamic but not predetermined input (Doraszelski and Jaumandreu, 2013; Ackerberg et al., 2015).

Restriction B of Assumption 4 - Timing of labor input choice: Labor is decided upon at time t or $t-i$ with $0 < i < 1$ and is, therefore, correlated with the productivity shock η_{bt}.

When labor is endogenous, it is common to instrument it with its lagged value, which needs to be taken into account when specifying the observed likelihood.

[20] The marginal model for the initial dependent variables is cluster-independent such that it cancels out from all posterior distributions specified during the estimation procedure.

Assumption 10: Labor demand parametrization: We specify a reduced-form equation for endogenous labor with exogenous instruments $k_{bt}, l_{bt-1}, \phi_{bt-1}$ and a normally distributed error term ($\zeta_{bt} \sim \mathcal{N}\left(0, (\sigma^s_\zeta)^2\right)$), such that:[21]

$$l_{bt} = \delta^s_0 + \delta_1 k_{bt} + \delta^s_2 \phi_{bt-1} + \delta^s_3 k_{bt-1} + \delta^s_4 l_{bt-1} + \zeta^s_{bt}. \tag{13}$$

Under the additional parametric assumptions imposed by Assumptions 8-10 and Restriction B of Assumption 4 on the timing of the labor input choice, we specify our Limited Information Maximum Likelihood (LIML) equation ($\mathcal{L}_B(\cdot)$), i.e., the partial log-likelihood to be maximized based on equation (10), as follows:

$$\mathcal{L}_B\left(\boldsymbol{\Theta}\right) = \sum_{b=1}^{B} log\left(\sum_{s=1}^{S} Pr_z(z_b = s|\omega_{b0}, k_{b0}, l_{b0}, ; \boldsymbol{\gamma}) \prod_{t=1}^{T} p^s_{\phi,l}(\phi_{bt}, l_{bt}|k_{bt}, \phi_{bt-1}, k_{bt-1}, l_{bt-1}; \boldsymbol{\theta}^s)\right). \tag{14}$$

Both the ML (12) and LIML approach (14) rely on the expectation-maximization algorithm. This algorithm allows us to jointly estimate the production function parameters, the cluster-specific parameters of the productivity process, and the unobserved cluster affiliation probabilities. We refer the interested reader to Online Appendix C for details on the estimation procedure.

3.2 Comparison with the GMM approach

The proposed (LI)ML approach in this paper introduces a novel estimation approach to account for latent heterogeneity in productivity evolution, achieved through the use of functional form restrictions. In comparison to the non-parametric GMM estimator, our estimation framework necessitates two to three parametric assumptions. We assume that productivity is log-normally distributed, the conditional cluster probability follows a multinomial logit model, and in the case that labor is not considered a predetermined input, we model endogenous labor as a reduced-form function incorporating exogenous instruments and a normally distributed error term. By examining the unconditional moment conditions associated with these approaches, we explore the strengths and weaknesses of the (LI)ML approach, as well as the restrictiveness of its parametric assumptions relative to the commonly used GMM approach. The just-identified moment conditions for a Cobb-Douglas production function with a linear Markov process for productivity are presented in Table 1.[22]

To begin, we examine how heterogeneity in the productivity process influences the afore-

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

[22] We refer to Online Appendix C for the derivation of these conditions.

Table 1: Overview of unconditional moment conditions

Approach	Moment condition
GMM	$E\left[\sum_{n=1}^{N}\mathbb{I}\left(E_b=n\right)\eta_{bt}^{n}\begin{pmatrix}k_{bt}\\ l_{bt(-1)}\\ 1\\ \omega_{bt-1}\end{pmatrix}\right]=0$
ML	$E\left[\sum_{s=1}^{S}Pr_z\left(z_b=s\vert\left\{k_{bt},l_{bt},\phi_{bt}\right\}_{t=0}^{T};\mathbf{\Theta}\right)\eta_{bt}^{s}\begin{pmatrix}\partial_{\beta_k}\eta_{bt}^{s}\\ \partial_{\beta_l}\eta_{bt}^{s}\\ 1\\ \omega_{bt-1}\end{pmatrix}\right]=0$
LIML	$E\left[\sum_{s=1}^{S}Pr_z\left(z_b=s\vert\left\{k_{b\tau},l_{bt},\phi_{bt}\right\}_{t=0}^{T};\mathbf{\Theta}\right)\left(\eta_{bt}^{s}-\frac{\sigma_{\eta,\zeta}^{s}}{\left(\sigma_{\zeta}^{s}\right)^2}\zeta_{bt}^{s}\right)\begin{pmatrix}\partial_{\beta_k}\eta_{bt}^{s}\\ \partial_{\beta_l}\eta_{bt}^{s}\\ 1\\ \omega_{bt-1}\end{pmatrix}\right]=0$

Note: The presented moment conditions represent the just-identified unconditional moment conditions for a Cobb-Douglas production function with a linear Markov process for productivity.

mentioned moment conditions. In the commonly employed GMM approach, heterogeneity is typically addressed through a *deterministic cluster affiliation* approach. Let us consider a scenario where we observe in the data a categorical variable E_b with N categories that determine cluster affiliation. This categorical variable governs the heterogeneity in the Markov process of productivity ($\sum_{n=1}^{N}\mathbb{I}\left(E_b=n\right)\eta_{bt}^{n}$ where $\mathbb{I}$ is an indicator function) and gives rise to the resulting moment conditions outlined in the first row of Table 1. It is important to note that researchers are typically uncertain whether this categorical variable accurately determines the actual cluster affiliation, and whether the number of categories corresponds to the number of clusters ($N=S$) or not. When $N<S$, the process becomes misspecified, the likelihood of which is present as discussed in the introduction. The production function is conventionally estimated with $N=1$.[23] Additionally, the presence of possible measurement error in the categorical variables contributes

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

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

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

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

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

to cluster misallocation, resulting in biased estimators.

In contrast, the proposed (LI)ML approach adopts a different strategy by modeling the unobserved cluster affiliation z_b as a random variable. The probability of this random cluster affiliation is determined solely based on readily available information, referred to as the *random cluster affiliation*, i.e., $\sum_{s=1}^{S} Pr_z \left(z_b = s | \{k_{bt}, l_{bt}, \phi_{bt}\}_{t=0}^{T} ; \mathbf{\Theta} \right) \eta_{bt}^{s}$. This approach is reflected in the second and third rows of Table 1. One key advantage of this approach is that it allows to estimate production function parameters without prior knowledge of firm cluster affiliation. In the unlikely circumstance where prior information regarding cluster affiliation is available, the proposed approach performs just as effectively as the deterministic GMM approach.[24] Moreover, by employing a random specification, this approach also accommodates measurement error in the categorical variables (as demonstrated in the Monte Carlo simulation below). For estimation purposes, however, we assume this conditional cluster probability follows a multinomial logit model.

Next, let's consider the instruments utilized in specifying the moment conditions. In the GMM approach, these instruments are typically the current levels of inputs, assuming labor is predetermined (k_{bt}, l_{bt}). However, when labor is not predetermined, it is common practice to instrument current labor with lagged labor values (k_{bt}, l_{bt-1}). In contrast, if labor is predetermined, we propose to employ a Maximum Likelihood (ML) estimation to estimate the production function. The resulting moment conditions (refer to Table 1, row two) rely on the derivatives of the error term with respect to input coefficients, denoted as $\partial_{\beta_x} \eta_{bt}$, rather than the input levels themselves, x_{bt}. In a scenario with a single cluster, these moment conditions are equivalent to those of a Nonlinear Least Square Estimator, and they do not necessitate additional parametric assumptions beyond those of the GMM approach.

If labor is not predetermined, we propose employing the LIML estimation approach to estimate the production function. In this approach, we adopt a control function framework where we model the endogeneity in the error term by incorporating a reduced-form specification for endogenous labor. Specifically, we express the error term as the difference between the latent normally distributed productivity shock, η_{bt}^{s}, and the weighted normally distributed error term from labor, $\left(\eta_{bt}^{s} - \frac{\sigma_{\eta,\zeta}^{s}}{\left(\sigma_{\zeta}^{s}\right)^2} \zeta_{bt}^{s} \right)$ (see Table 1, row three).

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

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

with the prior probabilities approximately equal to unity: $\hat{\gamma}_1^1 = \infty$ and $Pr_z(z_b = 1 | \mathbb{I}_b [s = 1] ; \hat{\boldsymbol{\gamma}}) \approx 1$. This prior information on cluster affiliation is validated by the data and results in a close to perfect identification of the posterior probability of cluster affiliation $\hat{z}_b^1 = Pr_z \left(z_b = 1 | \{k_{bt}, l_{bt}, \phi_{bt}\}_{t=0}^{T}, \mathbb{I}_b [s = 1] ; \hat{\mathbf{\Theta}} \right) \approx 1$.

Similar to ML, the LIML approach employs the derivatives of the error term with respect to input coefficients, $\partial_{\beta_x}\eta_{bt}$, as instruments instead of the input levels themselves, x_{bt}.

In Online Appendix C, we compare the empirical outcomes of this control function approach to address endogeneity issues with the traditional GMM approach, while keeping the heterogeneity in the productivity growth process and the instruments equal. Our findings reveal that a control function approach yields production function estimates highly comparable to those obtained using the non-parametric GMM approach. This empirical evaluation suggests that deviations between the GMM and (LI)ML estimators primarily arise from the distinct specifications of the instruments rather than from parametric restrictions.

3.3 Model selection

While the number of clusters S is assumed to be an exogenous variable in our economic model (see Section 2), we allow the data to determine this number. Testing the order of a finite mixture using likelihood ratio tests is difficult and rarely done, as regularity conditions that ensure a standard asymptotic distribution for the maximum likelihood estimates do not hold (Celeux et al., 2018). Therefore, we approach this step as a model selection problem, in which we estimate the model for several clusters and rely on evaluation criteria to determine the "true" number of clusters (Celeux et al., 2018). We rely on two evaluation criteria: the Bayesian information criterion (BIC) and the integrated complete-data likelihood Bayesian information criterion (ICLbic). If these evaluation criteria prefer a multi-cluster over a single-cluster model specification, we interpret this as a rejection of the homogeneity assumption for the productivity growth process.

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

$$BIC(S) = -2\mathcal{L}_{A,B}(\hat{\mathbf{\Theta}}) + np\, log(BT). \tag{15}$$

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

One limitation is that the BIC does not consider the purpose of the modeling. It does not account for the usefulness of additional clusters when assessing S, i.e., how well separated the different clusters are. Clusters are well separated if the posterior cluster probability $Pr_z\left(z_b = s|\left\{k_{bt}, l_{bt}, \phi_{bt}\right\}_{t=0}^{T}; \mathbf{\Theta}\right)$ is close to 1 for one component and close to 0 for all other components. Therefore, as an alternative criterion, we consider ICLbic, which selects S such that the resulting mixture model leads to a clustering of the data with the largest evidence base (Biernacki et al., 2000):

$$\begin{aligned} ICLbic(S) = -2\Bigg(\mathcal{L}_{A,B}(\hat{\mathbf{\Theta}}) + \Bigg[\sum_{s=1}^{S}\sum_{b=1}^{B}\sum_{t=1}^{T} Pr_z\left(z_b = s|\left\{k_{bt}, l_{bt}, \phi_{bt}\right\}_{t=0}^{T}; \hat{\mathbf{\Theta}}\right) \\ \times\, log\left(Pr_z\left(z_b = s|\left\{k_{bt}, l_{bt}, \phi_{bt}\right\}_{t=0}^{T}; \hat{\mathbf{\Theta}}\right)\right)\Bigg]\Bigg) + \frac{np}{2}log(BT). \end{aligned} \quad (16)$$

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

3.4 Monte Carlo

We conduct a Monte Carlo (MC) exercise to evaluate the estimator's performance. The focus is on the estimator's ability to recover unobserved heterogeneity in the productivity distribution. The setup of our MC analysis closely mimics that of Ackerberg et al. (2015), which builds on Syverson (2001) and Van Biesebroeck (2007). The key deviation from Ackerberg et al. (2015) is in the specification of the Markov process for productivity, which is assumed to differ between firm clusters.[25] Specifically, productivity is assumed to follow a finite mixture AR(1) process with two clusters ($S = 2$):

$$\omega_{bt} = \sum_{s=1}^{2} \mathbb{I}\left[z_b = s\right]\left[\alpha_0^s + \alpha_1^s \omega_{bt-1} + \eta_{bt}^s\right], \quad (17)$$

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

Firms make optimal capital investment choices to maximize the expected (discounted) value of future profits under convex capital adjustment costs such that the period t capital stock (K_{bt}) is determined by investment at $t-1$, i.e., $K_{bt} = (1-\delta)K_{bt-1} + I_{bt-1}$. Material inputs ($M_{bt}$) are chosen at t, while labor input (L_{bt}) is either assumed to be predetermined, and chosen at $t - i$, or not, and chosen at t (in the former case, labor

[25] See Online Appendix D for a complete description of the MC simulation.

is chosen with only knowledge of $e^{\omega_{b,t-i}}$ where $i \leq 1$, not $e^{\omega_{bt}}$) (Ackerberg et al., 2015). The production function is assumed Leontief in (and proportional to) materials, such that:

$$Y_{bt} = min\left\{K_{bt}^{\beta_k} L_{bt}^{\beta_l} e^{\omega_{bt}}, \beta_m M_{bt}\right\} e^{\varepsilon_{bt}}, \tag{18}$$

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

$$\frac{y_{bt}}{m_{bt}} = \beta_k k_{bt} + \beta_l l_{bt} + \omega_{bt} + \varepsilon_{bt}. \tag{19}$$

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

The second DGP (DGP2) introduces latent heterogeneity through differences in the productivity evolution between two clusters of firms which are observed by the researcher. 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.[26] The prior probability of cluster affiliation can be identified as:

$$Pr_z(z_b = s|k_{b0}, l_{b0}, \omega_{b0}; \boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2) = \frac{e^{\gamma_0^s + \gamma_k^s k_{b0} + \gamma_l^s l_{b0} + \gamma_\omega^s \omega_{b0} + \gamma_{cluster}^s \mathbb{I}(s=s)}}{\sum_{i=1}^{S} e^{\gamma_0^i + \gamma_k^i k_{b0} + \gamma_l^i l_{b0} + \gamma_\omega^i \omega_{b0} + \gamma_{cluster}^i \mathbb{I}(i=s)}}, \tag{20}$$

where the fact that cluster membership is identified by the researcher is captured by the indicator variable $\mathbb{I}(i = s)$. As such, the MC analysis on the DGP2 allows us to evaluate whether LIML can approximate the traditional GMM specification when heterogeneity is present but observed.

The third DGP (DGP3) builds on DGP2 but assumes that 10% of firms are misclassified in clusters as observed by the researcher. This is a more realistic scenario for researchers and allows a comparison of the deterministic approach and our proposed random approach to cluster affiliation. Finally, the fourth DGP (DGP4) increases the difficulty further by assuming heterogeneity is completely unobserved.

To estimate equation (19) for all DGPs, we follow the Ackerberg et al. (2015) estima-

[26] This is calculated as follows: $\frac{0.8}{1-0.77} - \frac{1}{1-0.7} \approx 0.145$

tion approach for the first stage. The second stage differentiates between identification strategies. Specifically, for each of the DGPs, we follow the discussion in Section 3.2 and, where possible, estimate: (i) a unitary cluster affiliation according to Ackerberg et al. (2015) (Uni. GMM); (ii) a deterministic cluster affiliation where the cluster identification variable is observed and accounted for in the Markov process (Det. GMM); and (iii) our proposed estimation approach with random one-cluster ($S = 1$) and two-cluster ($S = 2$) affiliation imposed, named as 1-comp. LIML and 2-comp. LIML, respectively.[27]

Table 2 displays the results of the MC analysis. Focusing on the evolution of productivity, we display the normalized, relative to the true parameter values, Mean Squared Error (NMSE)[28] of the Markov process parameters α_0, α_1, σ_η for both clusters, along with the NMSE of the average share-weighted productivity growth $\bar{\Omega} = \sum_{t=1}^{T} \sum_{b=1}^{B} \frac{share_{bt}\, \omega_{bt}}{T}$, where $share_{bt} = \frac{y_{bt}}{\sum_{b=1}^{B} y_{bt}}$, and of the average cluster productivity premium $\bar{\omega}^2 - \bar{\omega}^1$ where $\bar{\omega}^s = \sum_{b \in s} \frac{\omega_{bt}}{T \sum \mathbb{I}(b \in s)}$ for $s = 1, 2$. The NMSE allows to evaluate the bias and variance of the estimator in one statistic. It should be noted that in the case of the 2-comp. LIML, we rely on the model-identified rather than the imposed cluster affiliation to calculate these statistics.

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

The DGP2 reveals a bias in the productivity evolution parameters when cluster heterogeneity is present but not controlled for (see the Uni. GMM and the 1-comp. LIML). The Det. GMM and the 2-comp. LIML accurately control for this heterogeneity and showcase very similar performance. Note that the Det. GMM does not accurately identify differences in variance across components, which is essential for valid inference. The bias in the Markov process parameters translates to a strong bias in the average cluster productivity premium. The bias of the average share-weighted productivity growth is much smaller in magnitude.

When the analysis is based on mismeasured cluster affiliations (DGP3), the Det. GMM yields biased estimates. Only the 2-comp. LIML remains relatively robust. This behavior can be ascribed to the identification of cluster affiliation, which relies on the information available in the initial conditions, the evolution of productivity, and the cluster affiliation indicator. In DGP4, finally, the researcher is not able to rely on the Det. GMM approach

[27] For all estimators, starting values for the parameters are set to $\beta_k = 0.3$ and $\beta_l = 0.7$ while the true values are $\beta_k = 0.4$ and $\beta_l = 0.6$.

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

Table 2: Monte Carlo results

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

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

anymore, while the performance of the 2-comp. LIML approach remains relatively robust.

4 Application to firm-level data

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

We retain a set of active firms that report output, capital stock at the beginning of the year, number of employees in full-time equivalents (FTE), and material costs.[29] This database is combined with a rich set of firm-level characteristics considered relevant for productivity growth in the literature, including firm age, industry affiliation and whether or not the firm engages in export, import and FDI activities. The data is obtained from the Belgian Balance Sheet Transaction Trade Dataset and a Belgian survey on FDI.[30] All monetary variables are deflated using the appropriate industry-level deflators constructed from national accounts.

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

We parametrize the production function $f(\cdot;\boldsymbol{\beta})$ assuming both a gross-output (Gandhi et al., 2020) and value-added (Ackerberg et al., 2015) production function under both a Cobb-Douglas and Translog specification, with labor assumed not to be predetermined. These production functions are estimated using either a GMM estimation approach without allowing for unobserved heterogeneity in a linear Markov process $g(\omega_{bt-1},\boldsymbol{\alpha})$

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

[30] A similar database has already been used for productivity estimations by, among others, Mion and Zhu (2013) and Forlani et al. (2023).

[31] See Table A.1 in the Online Appendix for summary statistics of this sector and industries.

or using the proposed LIML with increasing heterogeneity in a linear Markov process $g^s(\omega_{bt-1}, \boldsymbol{\alpha}^s)$ (with the total number of clusters S limited to $S = 10$). For space considerations and conciseness, we discuss here the estimation results for a value-added Translog production function of a specific sector (i.e., industry 22) and refer the reader to Section 5 for a complete discussion of the estimation results for all remaining specifications and industries.

4.1 Production function estimates

Table 3 presents the production function estimates. The average output elasticities and returns to scale (RTS) shown in the table's first three rows indicate small, though not statistically significant, differences between the GMM and 1-comp. LIML. Most likely, these differences are linked to differences in efficiency. Specifically, the instruments for LIML are constructed optimally using the nonlinear model specification $E\left[\frac{\partial \phi_{bt}}{\partial \boldsymbol{\beta}}\eta_{bt}\right] = 0$. In contrast, the GMM typically relies on factor input and output levels to specify moment conditions (Ackerberg et al., 2015; Gandhi et al., 2020) (see also section 3.2).[32]

Interestingly, the production function estimates are robust to the relaxation of the homogeneity assumption concerning the productivity growth process. Comparing the output elasticities across models with increasing heterogeneity in the productivity process (1-comp. LIML up to 7-comp. LIML),[33] we observe that the point estimates are not identical as the number of clusters increases. However, they do not differ significantly statistically from the 1-comp. LIML estimates. We demonstrate in Online Appendix F that this robustness is not specific to the methodology used in this paper.

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

4.2 Latent heterogeneity in the productivity growth process

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

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

[33]We limit to exposition to 7 components in the paper for visual reasons. See Online Appendix A for the full results.

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

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

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

geneity in productivity growth. The goodness-of-fit indicators reported at the bottom of Table 3 demonstrate an increasingly good model fit as the number of clusters increases, with an optimal number of six clusters as indicated by the decrease in the BIC and ICL-bic up to the six-comp. LIML. These six clusters are well separated, as indicated by the posterior probabilities (see Online Appendix A, Figure A.1).

The cluster-specific coefficients for the evolution of productivity (in logs) are displayed in Table 4. We identify firm-cluster affiliation by choosing the cluster with the maximal posterior cluster affiliation probability per firm. We observe heterogeneity in both the constant (α_0^s) and auto-regressive parameters (α_1^s) of the productivity process across clusters, leading to a cluster hierarchy based on stationary average productivity levels (μ_ω^s). For instance, cluster 3 has a clear, although not significant, productivity advantage

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

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

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

over cluster 4, with a premium of around $13 - 12.8 \approx 20\%$.[34]

In addition, we observe significant heterogeneity in the volatility of the distribution of unexpected shocks to productivity (σ_η) and stationary volatility (σ_ω) that associate with stationary average productivity levels. Highly volatile productivity processes, such as those of clusters 1, 3, and 6, correlate with a relatively higher average productivity level. This indicates that firms that end up on the right tail of the productivity distribution have done so through a relatively volatile productivity growth process. However, high volatility in productivity does not mean that this volatility is equally persistent, as can be deduced from the auto-regressive parameters (α_1^s). Using an impulse response function approach in Figure 2, we demonstrate that an unexpected shock to productivity has a relatively less sizable long-lasting influence in a cluster with a volatile productivity growth process compared to one with a relatively more stable growth process.

4.3 Characterizing latent heterogeneity in productivity

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

[34]This distinction can also be visually evaluated based on the cluster-specific productivity densities displayed in Figure A.2 in Online Appendix A.

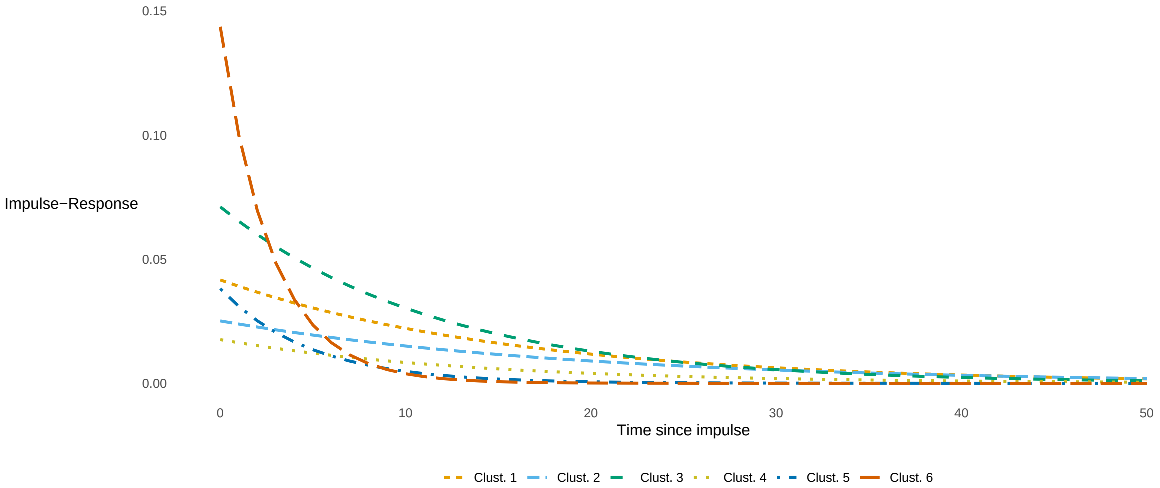

Figure 2: Impulse-response function of the cluster-specific productivity process.

Note: 50-year response function of the evolution of productivity of each cluster s after a one standard deviation unexpected productivity impulse: $IRF^s = \sum_{t=0}^{T=50} \sigma_\eta^s (\alpha_1^s)^t$. Cluster affiliation is determined as the maximal posterior cluster affiliation probability. The results shown are for estimates obtained using a panel of 626 firms and 4,399 observations in the Belgian NACE Rev. 22 industry for the period 2008-2018.

strength of the proposed estimator and correlate it with the estimated productivity of firm clusters.

The estimation results reported above rely on a base specification of cluster probabilities, conditioning only on initial capital, labor, and productivity (see equation (11)). This specification is derived under the assumption that the initial conditions contain sufficient information to identify cluster affiliation. If this assumption fails to hold, augmenting the base specification with additional, economically relevant (see the Introduction) firm-level characteristics is necessary to help improve the identification of cluster affiliation. To test this hypothesis, we augment equation (11) to the following multinomial logistic specification:

$$ln\frac{Pr_z(z_b = s|k_{b0}, l_{b0}, \omega_{b0}, \boldsymbol{e}_b; \boldsymbol{\gamma}^s)}{Pr_z(z_b = 1|k_{b0}, l_{b0}, \omega_{b0}, \boldsymbol{e}_b; \boldsymbol{\gamma}^1)} = \gamma_0^s + \gamma_1^s k_{b0} + \gamma_2^s l_{b0} + \gamma_3^s \omega_{b0} + \gamma_4^s age_{b0} + \gamma_5^s EXP_b + \gamma_6^s IMP_b + \gamma_7^s FDI_b, \quad \forall s = 2, \ldots, S \tag{21}$$

where cluster probabilities are specified conditional on initial capital, labor, and productivity as well as additional firm characteristics represented in the vector $\boldsymbol{e}_b = \{age_b, EXP_b, IMP_b, FDI_b\}$, such as initial age ($age_b$), and indicators of export (EXP_b), import (IMP_b) and FDI activity (FDI_b) over the sample period.[35] Furthermore, we specify a version of (21) without initial capital, labor, and productivity. If the considered firm-

[35] We classify firms as respectively exporters, importers, or engaged in FDI if they export, import, or engage in FDI at least at one point during the sample period. Firms that do not export, import, or engage in FDI over the entire sample period are classified as non-exporters, non-importers, or firms not engaged in FDI and are chosen as the reference group.

level characteristics contain sufficient information to group firms into clusters, we expect this specification to perform as well as our base specification.

We rely on the two augmented specifications discussed above to re-estimate the production function with 6 clusters. The resulting log-likelihood, BIC, and ICLbic are reported in Table 5. First, we focus on the differences between the base and augmented specifications and conclude that the former is preferred. The increase in log-likelihood obtained by the augmented specification is insufficient to warrant the increase in the number of parameters, as indicated by the smaller BIC and ICLbic indicators in absolute value relative to the base specification. The stability of the base specification to alternative specifications of cluster probabilities is in line with our Monte Carlo results and speaks to the ability of the estimator to identify firm clusters without additional information. Furthermore, this stability implies a substantial correlation between latent heterogeneity and initial conditions. Specifically, additional firm-level characteristics appear to have limited explanatory power once initial conditions are controlled for.

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

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

Notes: The base specification refers to equation (11), the augmented specification refers to equation (21), and the specification without initial capital, labor, and productivity refers to equation (21). BIC stands for Bayesian information criterion, and ICLbic for integrated complete-data likelihood Bayesian information criterion. Estimates are obtained from a 6-cluster value-added Translog production function with endogenous labor using a panel of 626 firms and 4,399 observations in the Belgian NACE Rev. 22 industry for 2008-2018.

To demonstrate how instrumental initial conditions are for cluster affiliation identification, 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. The BIC and ICLbic reaffirm the base specification's superior performance. Therefore, even when firm-level information regarding age and the internationalization status of a firm is available, a significant share of the heterogeneity in productivity remains latent and cannot be accounted for by these observable variables alone.

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

Table 6: Average cluster characteristics

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

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

Firm age, then, seems to be associated with the persistence of the productivity growth process. The clusters with relatively less persistent growth processes (clusters 3, 5, and 6) contain, on average, younger firms than those with relatively persistent growth (clusters 1, 2, and 4), which contain older firms on average. In addition, we observe that clusters 1 and 4 have a larger firm size in terms of initial output, capital, labor, and internationalization status, i.e., export, import, and FDI activity. Interestingly, there is a relatively large probability of importers belonging to clusters 1 and 3, which are relatively more volatile. This could be linked to the recent discussions about exposure to foreign shocks through global supply chains (Baldwin et al., 2022, 2023). However, since the productivity estimates combine efficiency and demand drivers, we do not engage in a more detailed analysis of the anatomy of these heterogeneous effects and instead focus on understanding their economic relevance.

4.4 The impact of latent heterogeneity on exporter productivity

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

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

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

$$\begin{aligned}
\Omega_t &= \sum_{b=1}^{B} share_{bt}\, \omega_{bt} \\
&= \frac{1}{G} \sum_{g=EXP,NONEXP} \Bigg[\bar{\omega}_t^g + \underbrace{\sum_{b=1}^{B} \left(share_{bt} - \overline{share}_t^g \right) (\omega_{bt} - \bar{\omega}_t^g)}_{\text{Within-group covariance}} \\
&\qquad\qquad + \underbrace{(share_t^g - \frac{1}{G})(\Omega_t^g - \bar{\Omega}_t)}_{\text{Between-group covariance}} \Bigg],
\end{aligned} \tag{22}$$

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

We calculate this decomposition of aggregate productivity for different productivity indices obtained from different estimation methodologies and different specifications of heterogeneity in productivity.[36] Specifically, we estimate productivity using the GMM and LIML identification strategies with (i) a *base* specification: $\omega_{bt} = \alpha_0 + \alpha_1 \omega_{bt-1} + \eta_{bt}$, (ii) a *deterministic* control for exporter status, $\omega_{bt} = \alpha_0 + \alpha_1 \omega_{bt-1} + \alpha_2 EXP_b + \alpha_3 \omega_{bt-1} EXP_b + \eta_{bt}$, and (iii) a more *exhaustive* set of controls for heterogeneity in productivity:

$$\begin{aligned} \omega_{bt} = \alpha_0 + \alpha_1 \omega_{bt-1} + \alpha_3 age_{b0} &+ \alpha_4 EXP_b + \alpha_5 \omega_{bt-1} EXP_b \\ &+ \alpha_6 IMP_b + \alpha_7 \omega_{bt-1} IMP_b \\ &+ \alpha_8 FDI_b + \alpha_9 \omega_{bt-1} FDI_b + \eta_{bt}. \end{aligned} \tag{23}$$

Similarly, we obtain productivity from the finite mixture LIML identification strategy with the optimal number of six clusters and (i) the *base* specification for cluster affiliation (11), (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 (21).

Figure 3 displays the evolution of the obtained aggregate productivities and their decomposition across estimation methodologies and specifications. Focusing on aggregate productivity (the left column), we observe that the evolution over time is very similar across estimation methodologies and specifications. This behavior can be attributed to the robustness of the production function and productivity estimates to the homogeneity assumption of the evolution of productivity, as reported above. This is also in line with the MC analysis (see subsection 3.4), based on which we expect a strong bias in the productivity premium but a smaller bias on the average share-weighted productivity growth.

Exploring the decomposition of this aggregate productivity, we observe differences depending on the estimation methodology and the specification of heterogeneity in the Markov process, for both the GMM and LIML estimation methodologies. This contrasts with the robustness of the finite mixture LIML across specifications. For instance, the export premium—the difference between the average productivity of exporters (dashed line) and non-exporters (continuous line) in the second column of Figure 3— evolves from 1.97%, for the base specification, to 3.16% for the deterministic and 2.16% for the exhaustive specification of heterogeneity for the LIML methodology (see also Online Appendix Table A.2). Notably, the observed differences in export premia between heterogeneity specifications are relatively constant over time. In comparison, the export premium is approximately 1.6% for all three specifications of the finite mixture LIML.

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

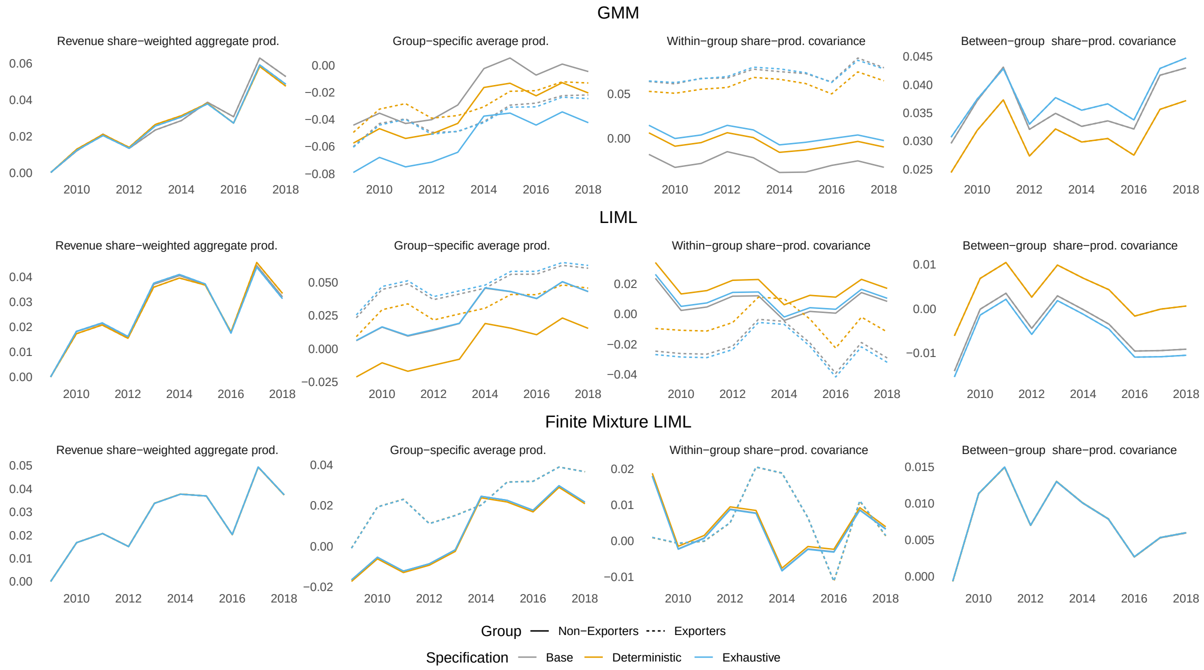

Figure 3: Evolution of aggregate productivity and its decomposition for exporting- and non-exporting firms.

Notes: GMM, LIML, and finite mixture LIML refer to the productivity estimation methodologies, while Base, Deterministic, and Exhaustive refer to the specification of heterogeneity within these methodologies, i.e., see equations (23) and (21).

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

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

5 Robustness

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

It could be of concern that our results are specific to the Belgian firm-level dataset. Therefore, we also apply the analysis to the Chilean firm-level dataset used by Gandhi et al. (2020). Results in Online Appendix E reaffirm the findings obtained with the Belgian dataset. We reaffirm the robustness of the production function coefficients to the relaxation of the homogeneity assumption in productivity, the presence of heterogeneity in the production process, the sufficiency of the initial conditions to identify cluster affiliation, and the accuracy of the proposed methodology to identify productivity differences between groups of firms and their respective contribution to aggregate productivity growth. Moreover, we find that the impact of relaxing the homogeneity assumption in the evolution of productivity on the calculation of export premia is more pronounced for Chile than for Belgium, with differences in the estimated productivity premia of up to 30%.

6 Conclusion

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

Of interest to the applied economist, production function coefficients and the overall evolution of aggregate productivity obtained from current productivity estimation methodologies seem robust to this latent heterogeneity. In line with Fernandes (2007) and De Loecker (2013), however, we find that the characterization of heterogeneity in productivity is not robust to latent heterogeneity. Additional explanatory variables expected to capture differences in the evolution of productivity, such as the firm's export, import, and FDI status, are associated with multiple productivity clusters obtained from the proposed method. This indicates heterogeneity in productivity beyond what is captured by the observed firm-level characteristics. As a result, current productivity estimation methodologies require additional firm-level information that remains notoriously unavailable, especially for hard-to-quantify productivity determinants such as intangible capital or managerial capacity. The proposed methodology, on the other hand, maintains its performance irrespective of the presence of this type of supplementary information.

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

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

References

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

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

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

Ackerberg, D. A., J. Hahn, and Q. Pan (2022). Nonparametric Identification Using Timing and Information Set Assumptions with an Application to Non-Hicks Neutral Productivity Shocks. Mimeo.

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.

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

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

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

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

Baldwin, R., R. Freeman, and A. Theodorakopoulos (2022, September). Horses for Courses: Measuring Foreign Supply Chain Exposure.

Baldwin, R., R. Freeman, and A. Theodorakopoulos (2023, October). Hidden Exposure: Measuring US Supply Chain Reliance.

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Fernandes, A. M. (2007). Trade policy, trade volumes and plant-level productivity in colombian manufacturing industries. Journal of International Economics 71(1), 52–71.

Forlani, E., R. Martin, G. Mion, and M. Muûls (2023, August). Unraveling Firms: Demand, Productivity and Markups Heterogeneity. The Economic Journal 133(654), 2251–2302.

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

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

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

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

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

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

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

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

Higgins, A. and K. Jochmans (2023a, November). Identification of mixtures of dynamic discrete choices. Journal of Econometrics 237(1), 105462.

Higgins, A. and K. Jochmans (2023b). LEARNING MARKOV PROCESSES WITH LATENT VARIABLES FROM LONGITUDINAL DATA.

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.

Hu, Y. and M. Shum (2012, November). Nonparametric identification of dynamic models with unobserved state variables. Journal of Econometrics 171(1), 32–44.

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

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

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

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

Kasahara, H., P. Schrimpf, and M. Suzuki (2023, May). Identification and Estimation of Production Function with Unobserved Heterogeneity.

Kasahara, H. and K. Shimotsu (2009). Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices. Econometrica 77(1), 135–175.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Rubens, M. (2021). Market structure, oligopsony power, and productivity. Technical report, mimeo.

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.

Tybout, J. R. (1992, July). Making noisy data sing: Estimating production technologies in developing countries. Journal of Econometrics 53(1), 25–44.

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

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

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

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

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

Online Appendix
Identifying Latent Heterogeneity in Productivity

Ruben Dewitte, Catherine Fuss, and Angelos Theodorakopoulos*

January, 2024 [Click here for latest version]

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

Appendix A Additional Figures and Tables

A.1 Figures

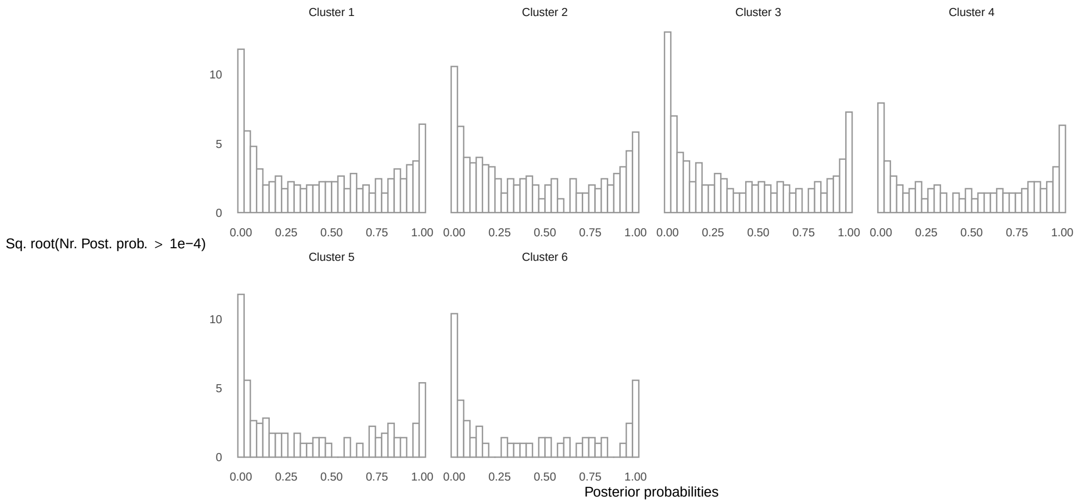

Figure A.1: Histogram of posterior probabilities for a 6-cluster (ACF) value-added Translog production function of NACE Rev. 22 estimated with LIML.

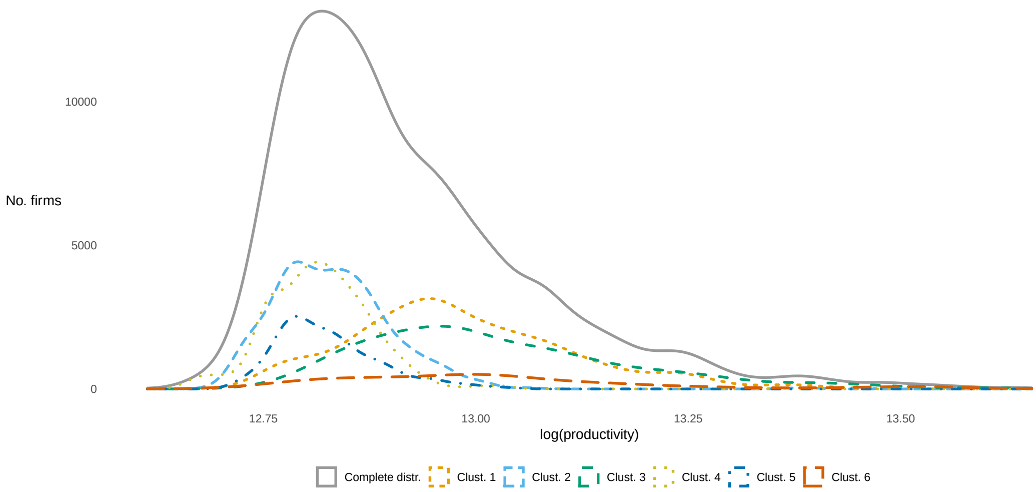

Figure A.2: Complete and cluster-specific density of productivity in 2013 obtained from a 6-cluster value-added Translog production function of NACE Rev. 22 estimated with LIML.

A.2 Tables

Table A.1: Summary Statistics

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

Table A.2: Average export premia across productivity estimation methodologies and specifications

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

Notes: Export premia obtained from a log-linear regression with year dummies t, $\omega_{bt} = \alpha Exp_b + t + \epsilon_{bt}$, where productivity is obtained from GMM, LIML and Finite Mixture LIML estimation methodologies with a base, deterministic or exhaustive specification of heterogeneity within these methodologies, i.e. see equations (23) and (21). Standard errors between brackets are obtained from the OLS regression.

Appendix B Nonparametric Identification

The equation below rewrites the joint data distribution to a structure that allows nonparametric identification of its elements.

$$
\begin{aligned}
&p_{\phi,k,l,z}\left(\left\{\{\phi_{bt},k_{bt},l_{bt}\}_{t=0}^{T},z_b\right\}_{b=1}^{B}\right)\\
&\quad=\prod_{b=1}^{B}\sum_{s=1}^{S}p_{\phi_{b0},k_{b0},l_{b0},z_b}\left(\phi_{b0},k_{b0},l_{b0},z_b\right)\\
&\qquad\qquad\times\prod_{t=1}^{T}p_{\phi,k,l,z}\left(\phi_{bt},k_{bt},l_{bt},z_b=s|\left\{\phi_{b\tau},k_{b\tau},l_{b\tau}\right\}_{\tau=1}^{t-1},z_b=s\right)\\
&\quad=\prod_{b=1}^{B}\sum_{s=1}^{S}Pr_z\left(z_b=s|\phi_{b0},k_{b0},l_{b0}\right)p_{\phi_{b0},k_{b0},l_{b0}}\left(\phi_{b0},k_{b0},l_{b0}\right)\\
&\qquad\qquad\times\prod_{t=1}^{T}p^{s}_{\phi,k,l,z}\left(\phi_{bt},k_{bt},l_{bt}|\left\{\phi_{b\tau},k_{b\tau},l_{b\tau}\right\}_{\tau=1}^{t-1}\right)\\
&\quad=\prod_{b=1}^{B}\sum_{s=1}^{S}Pr_z\left(z_b=s|\phi_{b0},k_{b0},l_{b0}\right)p_{\phi_{b0},k_{b0},l_{b0}}\left(\phi_{b0},k_{b0},l_{b0}\right)\\
&\qquad\qquad\times\prod_{t=1}^{T}p^{s}_{\phi}\left(\phi_{bt}|k_{bt},l_{bt},\left\{\phi_{b\tau},k_{b\tau},l_{b\tau}\right\}_{\tau=1}^{t-1}\right)P^{s}_{k,l}\left(k_{bt},l_{bt}|\left\{\phi_{b\tau},k_{b\tau},l_{b\tau}\right\}_{\tau=1}^{t-1}\right)\\
&\quad=\prod_{b=1}^{B}\sum_{s=1}^{S}Pr_z\left(z_b=s|\omega_{b0},k_{b0},l_{b0}\right)p_{\phi_{b0},k_{b0},l_{b0}}\left(\phi_{b0},k_{b0},l_{b0}\right)\\
&\qquad\qquad\times\prod_{t=1}^{T}p^{s}_{\phi}\left(\phi_{bt}|k_{bt},l_{bt},\phi_{bt-1},k_{bt-1},l_{bt-1}\right)p^{s}_{k,l}\left(k_{bt},l_{bt}|\omega_{bt-1},k_{bt-1},l_{bt-}\right)\\
&\quad=\prod_{b=1}^{B}\sum_{s=1}^{S}Pr_z\left(z_b=s|\omega_{b0},k_{b0},l_{b0}\right)p_{\phi_{b0},k_{b0},l_{b0}}\left(\phi_{b0},k_{b0},l_{b0}\right)\\
&\qquad\qquad\times\prod_{t=1}^{T}p^{s}_{\phi,k,l}\left(\phi_{bt},k_{bt},l_{bt}|\phi_{bt-1},k_{bt-1},l_{bt-1}\right)
\end{aligned}
\tag{B.1}
$$

The first equality follows from the assumption that firm types are determined in the initial period (Assumption 1). The second equality factorizes the joint distribution into a model for z_b conditional on the initial information set and a marginal model for this initial information set. Equality three restricts the type-specific conditional distribution function for the observable information set to a first-order Markov process based on the regression specification for ϕ_{bt} (eq. (9)) and on the assumptions on the timing of the input choices (Assumption 4). Notice that we rely on the functional relation between ϕ_{bt} and ω_{bt}, given k_{bt} and l_{bt} in equality 4 and 5 to swap conditioning variables.

Appendix C Estimation

This section describes the production function estimation techniques used in this paper. We summarize the proxy variable Ackerberg et al. (2015) and the first-order condition Gandhi et al. (2020) methods for the case when $S = 1$, before advancing to our proposed Mixture (LI)ML estimator for $S \geq 1$.

C.1 Proxy variable methods

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

$$\frac{y_{bt}}{m_{bt}} = f_{kl}\left(k_{bt}, l_{bt}\right) + \mathbb{M}^{-1}\left(k_{bt}, l_{bt}, m_{bt}\right) + \varepsilon_{bt}. \tag{C.1}$$

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

$$\phi_{bt} \equiv \frac{y_{bt}}{m_{bt}} - \varepsilon_{bt} = f_{kl}\left(k_{bt}, l_{bt}\right) + \omega_{bt}. \tag{C.2}$$

Building on the assumption that productivity evolves according to a first-order Markov process (Assumption 4), we obtain the *second stage* estimation equation:

$$\phi_{bt} = f_{kl}\left(k_{bt}, l_{bt}\right) + g\left(\phi_{bt-1} - f_{kl}\left(k_{bt-1}, l_{bt-1}\right)\right) + \eta_{bt}. \tag{C.3}$$

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:

$$E\left[\eta_{bt} | k_{bt}, l_{bt(-1)}, \phi_{bt-1}\right] = 0. \tag{C.4}$$

We parametrize equation (C.3) with production function coefficients $\boldsymbol{\beta}$ and specify a linear first-order Markov process $g(\phi_{bt-1} - f_{kl}(l_{bt-1}, k_{bt-1}; \boldsymbol{\beta}); \boldsymbol{\alpha}) = \boldsymbol{W}_{bt-1}\boldsymbol{\alpha}$ with $\boldsymbol{W}_{bt-1} = [1, \omega_{bt-1}]$, and $\theta = \{\boldsymbol{\beta}, \boldsymbol{\alpha}\}$ such that

$$\phi_{bt} = f_{kl}\left(k_{bt}, l_{bt}; \boldsymbol{\beta}\right) + \boldsymbol{W}_{bt-1}\boldsymbol{\alpha} + \eta_{bt}. \tag{C.5}$$

We transform the conditional moments specified above into unconditional moment conditions for actual estimation:

$$E\left[\eta_{bt}\begin{pmatrix}k_{bt}\\ l_{bt(-1)}\\ 1\\ \omega_{bt-1}\end{pmatrix}\right]=0. \tag{C.6}$$

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

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

$$\arg\min_{\boldsymbol{\beta}}\Lambda(\boldsymbol{\beta})=\arg\min_{\boldsymbol{\beta}}\left(\frac{\sum_{b=1}^B\sum_{t=1}^T\boldsymbol{Z}_{bt}^T\eta_{bt}}{B}\right)^T\left(\frac{\sum_{b=1}^B\sum_{t=1}^T\boldsymbol{Z}_{bt}^T\boldsymbol{Z}_{bt}}{B}\right)^{-1}\left(\frac{\sum_{b=1}^B\sum_{t=1}^T\boldsymbol{Z}_{bt}^T\eta_{bt}}{B}\right), \tag{C.7}$$

with the corresponding First-Order Condition (FOC):

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

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

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

$$\begin{aligned}&\arg\min_{\boldsymbol{\alpha}}\Lambda(\boldsymbol{\alpha})=\\ &\arg\min_{\boldsymbol{\alpha}}\left(\frac{\sum_{b=1}^B\sum_{t=1}^T\boldsymbol{W}_{bt}^T\eta_{bt}(\hat{\boldsymbol{\beta}})}{B}\right)^T\left(\frac{\sum_{b=1}^B\sum_{t=1}^T\boldsymbol{W}_{bt}^T\boldsymbol{W}_{bt}}{B}\right)^{-1}\left(\frac{\sum_{b=1}^B\sum_{t=1}^T\boldsymbol{W}_{bt}^T\eta_{bt}(\hat{\boldsymbol{\beta}})}{B}\right),\end{aligned} \tag{C.8}$$

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

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

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

C.2 First-order condition methods

As in the previous section, the estimation strategy proposed by Gandhi et al. (2020) consists of two stages, but for for a gross output production function. In the *first stage*, one relies on the materials as a proxy for productivity to single out the ex-post Hicks-neutral productivity shock and possible classical measurement error ε_{bt}:

Starting from a gross output production function:

$$y_{bt} = f^{klm}\left(k_{bt}, l_{bt}, m_{bt}\right) + \omega_{bt} + \varepsilon_{bt}, \tag{C.9}$$

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

$$log\left(\frac{P_t^M M_{bt}}{P_t^Y Y_{bt}}\right) = log\left(\mathcal{E}\right) + log\left(\frac{\partial f^{klm}\left(k_{bt}, l_{bt}, m_{bt}\right)}{\partial m_{bt}}\right) - \varepsilon_{bt} \tag{C.10}$$

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

$$\phi_{bt} = y_{bt} - \varepsilon_{bt} - \int \frac{\partial f^{klm}\left(k_{bt}, l_{bt}, m_{bt}\right)}{\partial m_{bt}} dm_{bt} = h_{kl}\left(k_{bt}, l_{bt}\right) + \omega_{bt}. \tag{C.11}$$

Relying on the productivity evolving according to a first-order Markov process, (Assumption 4), this results in the *second stage* estimation equation

$$\phi_{bt} = h_{kl}\left(k_{bt}, l_{bt}\right) + g\left(\phi_{bt-1} - h_{kl}\left(k_{bt-1}, l_{bt-1}\right)\right) + \eta_{bt}. \tag{C.12}$$

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:

$$E\left[\eta_{bt}|k_{bt}, l_{bt(-1)}, \phi_{bt-1}\right] = 0. \tag{C.13}$$

We parametrize equation (C.12) with production function coefficients $\boldsymbol{\beta}$ and specify a linear first-order Markov process $g(\phi_{bt-1} - f_{kl}(l_{bt-1}, k_{bt-1}; \boldsymbol{\beta}); \boldsymbol{\alpha}) = \boldsymbol{W}_{bt-1}\boldsymbol{\alpha}$ with $\boldsymbol{W}_{bt-1} = [1, \omega_{bt-1}]$, and $\theta = \{\boldsymbol{\beta}, \boldsymbol{\alpha}\}$ such that

$$\phi_{bt} = h_{kl}\left(k_{bt}, l_{bt}; \boldsymbol{\beta}\right) + \boldsymbol{W}_{bt-1}\boldsymbol{\alpha} + \eta_{bt}. \tag{C.14}$$

This specification takes a very similar form to the estimation equation for the proxy variable method specified above. The remaining unconditional moment conditions and optimization criteria are equivalent to those of the proxy variable methods expressed above (see equations (C.20), (C.7), and (C.8)).

C.3 Mixture (limited information) maximum likelihood

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

C.3.1 Estimation procedure

We rely on the expectation-maximization algorithm (McLachlan and Peel, 2000; Miljkovic and Grün, 2016) to maximize the ML specification (eq. (12)) or LIML specification (eq. 14). In the EM framework, the log-likelihoods specified in the main text are regarded as being incomplete as they do not account for the unobserved cluster affiliation. When accounting for cluster affiliation ($z_b^s = \mathbb{I}\left(z_b = s\right)$), the complete ($c$) partial log-likelihood specifications become:

$$\mathcal{L}_A^c(\boldsymbol{\Theta}, \boldsymbol{z}) = \sum_{b=1}^{B}\sum_{s=1}^{S} z_b^s log\left(Pr_z(z_b = s|\omega_{b0}, k_{b0}, l_{b0}, ; \boldsymbol{\gamma}) \prod_{t=1}^{T} p_\phi^s(\phi_{bt}|l_{bt}, k_{bt}, \phi_{bt-1}, k_{bt-1}, l_{bt-1}; \boldsymbol{\theta}^s)\right) \tag{C.15}$$

$$\mathcal{L}_B^c(\boldsymbol{\Theta}, \boldsymbol{z}) = \sum_{b=1}^{B}\sum_{s=1}^{S} z_b^s log\left(Pr_z(z_b = s|\omega_{b0}, k_{b0}, l_{b0}, ; \boldsymbol{\gamma}) \prod_{t=1}^{T} p_{\phi,l}^s(\phi_{bt}, l_{bt}|k_{bt}, \phi_{bt-1}, k_{bt-1}, l_{bt-1}; \boldsymbol{\theta}^s)\right), \tag{C.16}$$

which form the basis for our estimation procedure. We discuss first the more general LIML estimation approach when labor is decided upon at time t or $t-i$ with $0 < i < 1$, before discussing the ML approach where labor is decided upon at time $t-1$ and, therefore, predetermined.

C.3.2 Labor is decided upon at time t or $t-i$ with $0 < i < 1$

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

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

$$\begin{aligned}\hat{z}_b^s &= Pr_z\left(z_b = s|\left\{k_{bt}, l_{bt}, \phi_{bt}\right\}_{t=0}^T; (\boldsymbol{\Theta})^j\right)\\ &= \frac{Pr_z(z_b = s|\omega_{b0}, k_{b0}, l_{b0}; (\boldsymbol{\gamma}^s)^j) \prod_{t=1}^T p_{\phi,l}^s(\phi_{bt}, l_{bt}|k_{bt}, \phi_{bt-1}, l_{bt-1}, k_{bt-1}; \boldsymbol{\theta}^s)}{p_{\phi,l}\left(\left\{\phi_{bt}\right\}_{t=0}^T, \left\{l_{bt}\right\}_{t=0}^T; (\boldsymbol{\Theta})^j\right)}.\end{aligned}$$

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

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

$$\text{(i)} \quad \begin{aligned} \max_{(\boldsymbol{\theta})^{j+1}} \Lambda(\boldsymbol{\theta}^{j+1}) = \max_{(\boldsymbol{\theta})^{j+1}} \sum_{b=1}^{B}\sum_{s=1}^{S} Pr_z\left(z_b = s|\left\{k_{bt}, l_{bt}, \phi_{bt}\right\}_{t=0}^{T}; (\boldsymbol{\Theta})^j\right) \\ \times\, log\left(\prod_{t=1}^{T} p^s_{\phi,l}(\phi_{bt}, l_{bt}|k_{bt}, \phi_{bt-1}, k_{bt-1}, l_{bt-1}; (\boldsymbol{\theta}^s))^{j+1}\right); \end{aligned}$$

$$\text{(ii)} \quad \begin{aligned} \max_{(\boldsymbol{\gamma}^s)^{j+1}} \Lambda((\boldsymbol{\gamma}^s)^{j+1}) = \max_{(\boldsymbol{\gamma}^s)^{j+1}} \sum_{b=1}^{B}\sum_{s=1}^{S} Pr_z\left(z_b = s|\left\{k_{bt}, l_{bt}, \phi_{bt}\right\}_{t=0}^{T}; (\boldsymbol{\Theta})^j\right) \\ \times\, log\left(Pr_z(z_b = s|k_{b0}, l_{b0}, \omega_{b0}; (\boldsymbol{\gamma}^s)^{j+1})\right). \end{aligned}$$

This iterative process continues until there is relative stability between iterations j and $j+1$ in terms of the log-likelihood.

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

Following from Assumptions 8 and 10, the observed likelihood attains a bivariate normal specification:

$$p^s_{\phi,l}(\phi_{bt}, l_{bt}|k_{bt}, \phi_{bt-1}, k_{bt-1}, l_{bt-1}; \underbrace{\boldsymbol{\beta}, \boldsymbol{\alpha}^s, \boldsymbol{\delta}^s, \boldsymbol{\Sigma}^s}_{\equiv\boldsymbol{\theta}^s}) = \frac{e^{-\frac{1}{2}(\boldsymbol{\epsilon}^s)^T(\boldsymbol{\Sigma}^s)^{-1}(\boldsymbol{\epsilon}^s)}}{\sqrt{(2\pi)^2|\boldsymbol{\Sigma}^s|}}, \tag{C.17}$$

where $\boldsymbol{\epsilon}^s = \begin{bmatrix} \phi_{bt} - f^{kl}\left(k_{bt}, l_{bt}; \boldsymbol{\beta}\right) - g(\phi_{bt-1} - f^{kl}(k_{bt-1}, l_{bt-1}; \boldsymbol{\beta}); \boldsymbol{\alpha}^s) \\ l_{bt} - \delta^s_0 - \delta_1 k_{bt} - \delta^s_2\phi_{bt-1} - \delta^s_3 k_{bt-1} - \delta^s_4 l_{bt-1} \end{bmatrix}$ and $\boldsymbol{\Sigma}^s = \begin{bmatrix} (\sigma^s_\eta)^2 & \sigma^s_{\eta,\zeta} \\ \sigma^s_{\eta,\zeta} & (\sigma^s_\zeta)^2 \end{bmatrix}$.

To simplify the estimation procedure, we rely on the observation that equation (C.17) can be factorized into a density of the endogenous variables conditional on the instrumental variables, $p^s_{\phi,l}(\phi_{bt}, l_{bt}) = p^s_\phi(\phi_{bt}|l_{bt})p^s_l(l_{bt})$, such that

$$p^s_\phi(\phi_{bt}|k_{bt}, l_{bt}, \phi_{bt-1}, k_{bt-1}, l_{bt-1}; \boldsymbol{\beta}, \boldsymbol{\alpha}^s, \sigma^s_\eta, \sigma^s_{\eta,\zeta}) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\left[\left(\sigma^s_\eta\right)^2 - \frac{(\sigma^s_{\eta,\zeta})^2}{\left(\hat{\sigma}^s_\zeta\right)^2}\right]}} e^{-\frac{1}{2}\frac{\left(\eta^s_{bt} - \frac{\sigma^s_{\eta,\zeta}}{\left(\hat{\sigma}^s_\zeta\right)^2}\zeta^s_{bt}\right)^2}{(\sigma^s_\eta)^2 - \frac{\sigma^s_{\eta,\zeta}}{\left(\hat{\sigma}^s_\zeta\right)^2}}}, \tag{C.18}$$

and

$$p^s_l(l_{bt}|k_{bt}, \phi_{bt-1}, k_{bt-1}, l_{bt-1}; \boldsymbol{\delta}^s, \sigma^s_\zeta) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\left(\sigma^s_\zeta\right)^2}} e^{-\frac{1}{2}\left(\frac{\zeta^s_{bt}}{\sigma^s_\zeta}\right)^2}. \tag{C.19}$$

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

1. $$\nabla_{\boldsymbol{\delta}^s}\Lambda(\boldsymbol{\theta}) = 0 = -\frac{1}{\left(\sigma^s_\zeta\right)^2}\sum_{b=1}^{B} Pr_z\left(z_b = s|\left\{k_{bt}, l_{bt}, \phi_{bt}\right\}_{t=0}^{T}; \boldsymbol{\Theta}\right)\sum_{t=1}^{T}\boldsymbol{Z}_{bt}^{T}\left(l_{bt} - \boldsymbol{Z}_{bt}\boldsymbol{\delta}^s\right)$$
$$\Leftrightarrow \boldsymbol{\delta}^s = \left(\sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{Z}_{bt}^{T} Pr_z\left(z_b = s|\left\{k_{bt}, l_{bt}, \phi_{bt}\right\}_{t=0}^{T}; \boldsymbol{\Theta}\right)\boldsymbol{Z}_{bt}\right)^{-1}$$
$$\times\sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{Z}_{bt}^{T} Pr_z\left(z_b = s|\left\{k_{bt}, l_{bt}, \phi_{bt}\right\}_{t=0}^{T}; \boldsymbol{\Theta}\right) l_{bt};$$

2. $$\left(\hat{\sigma}^s_\zeta\right)^2 = \frac{\sum_{b=1}^{B}\sum_{t=1}^{T} Pr_z\left(z_b = s|\left\{k_{bt}, l_{bt}, \phi_{bt}\right\}_{t=0}^{T}; \boldsymbol{\Theta}\right)\left(\zeta^s_{bt}\right)^2}{\sum_{b=1}^{B}\sum_{t=1}^{T} Pr_z\left(z_b = s|\left\{k_{bt}, l_{bt}, \phi_{bt}\right\}_{t=0}^{T}; \boldsymbol{\Theta}\right)};$$

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

3. $$\nabla_{\boldsymbol{\alpha}^s}\Lambda(\boldsymbol{\theta}) = 0$$
$$= -\frac{1}{\left(\sigma^s_\eta\right)^2 - \frac{\left(\sigma^s_{\eta,\zeta}\right)^2}{\left(\sigma^s_\zeta\right)^2}}\sum_{b=1}^{B} Pr_z\left(z_b = s|\left\{k_{bt}, l_{bt}, \phi_{bt}\right\}_{t=0}^{T}; \boldsymbol{\Theta}\right)$$
$$\times\sum_{t=1}^{T}\left(\nabla_{\boldsymbol{\alpha}^s}\left(\eta^s_{bt} - \frac{\sigma^s_{\eta,\zeta}}{\left(\sigma^s_\zeta\right)^2}\zeta^s_{bt}\right)\right)^T\left(\eta^s_{bt} - \frac{\sigma^s_{\eta,\zeta}}{\left(\sigma^s_\zeta\right)^2}\zeta^s_{bt}\right)$$
$$\Leftrightarrow \boldsymbol{\alpha}^s = \left(\sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{W}_{bt-1}^{T} Pr_z\left(z_b = s|\left\{k_{bt}, l_{bt}, \phi_{bt}\right\}_{t=0}^{T}; \boldsymbol{\Theta}\right)\left(\boldsymbol{W}_{bt-1} - \frac{\sigma^s_{\boldsymbol{W}_{bt-1},\zeta}}{\left(\sigma^s_\zeta\right)^2}\zeta^s_{bt}\right)\right)^{-1}$$
$$\times\sum_{b=1}^{B}\sum_{t=1}^{T}\boldsymbol{W}_{bt-1}^{T} Pr_z\left(z_b = s|\left\{k_{bt}, l_{bt}, \phi_{bt}\right\}_{t=0}^{T}; \boldsymbol{\Theta}\right)\left(\omega_{bt} - \frac{\sigma^s_{\omega,\zeta}}{\left(\sigma^s_\zeta\right)^2}\zeta^s_{bt}\right)$$

4. $\nabla_{\boldsymbol{\beta}}\Lambda\left(\boldsymbol{\theta}\right)=0=\sum_{s=1}^{S}-\frac{1}{\left(\sigma_{\eta}^{s}\right)^{2}-\frac{(\sigma_{\eta,\zeta}^{s})^{2}}{\left(\sigma_{\zeta}^{s}\right)^{2}}}\sum_{b=1}^{B}Pr_{z}\left(z_{b}=s|\left\{k_{bt},l_{bt},\phi_{bt}\right\}_{t=0}^{T};\boldsymbol{\Theta}\right)$

$$\times\sum_{t=1}^{T}\left(\nabla_{\boldsymbol{\beta}}\eta_{bt}^{s}\right)^{T}\left(\eta_{bt}^{s}-\frac{\sigma_{\eta,\zeta}^{s}}{\left(\sigma_{\zeta}^{s}\right)^{2}}\zeta_{bt}^{s}\right)$$

$$\Leftrightarrow 0=\sum_{s=1}^{S}\sum_{b=1}^{B}\sum_{t=1}^{T}\frac{1}{\left(\sigma_{\eta}^{s}\right)^{2}-\frac{\sigma_{\eta,\zeta}^{s}}{\left(\sigma_{\zeta}^{s}\right)^{2}}}Pr_{z}\left(z_{b}=s|\left\{k_{bt},l_{bt},\phi_{bt}\right\}_{t=0}^{T};\boldsymbol{\Theta}\right)$$

$$\times\left(\nabla_{\boldsymbol{\beta}}(\eta_{bt}^{s})\right)^{T}\left(\eta_{bt}^{s}-\frac{\sigma_{\eta,\zeta}^{s}}{\left(\sigma_{\zeta}^{s}\right)^{2}}\zeta_{bt}^{s}\right),$$

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. $\left(\hat{\sigma}_{\eta}^{s}\right)^{2}=\frac{\sum_{b=1}^{B}\sum_{t=1}^{T}Pr_{z}\left(z_{b}=s|\left\{k_{bt},l_{bt},\phi_{bt}\right\}_{t=0}^{T};\boldsymbol{\Theta}\right)\left(\eta_{bt}^{s}\right)^{2}}{\sum_{b=1}^{B}\sum_{t=1}^{T}Pr_{z}\left(z_{b}=s|\left\{k_{bt},l_{bt},\phi_{bt}\right\}_{t=0}^{T};\boldsymbol{\Theta}\right)}$;

6. $\hat{\sigma}_{\eta,\zeta}^{s}=\frac{\sum_{b=1}^{B}\sum_{t=1}^{T}Pr_{z}\left(z_{b}=s|\left\{k_{bt},l_{bt},\phi_{bt}\right\}_{t=0}^{T};\boldsymbol{\Theta}\right)\eta_{bt}^{s}\zeta_{bt}^{s}}{\sum_{b=1}^{B}\sum_{t=1}^{T}Pr_{z}\left(z_{b}=s|\left\{k_{bt},l_{bt},\phi_{bt}\right\}_{t=0}^{T};\boldsymbol{\Theta}\right)}$.

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

For comparison with the GMM estimation methodology specified above, we write down unconditional moment conditions related to the second-stage log-likelihood estimation. For a Cobb-Douglas production function with a linear Markov process for productivity, the set of four second-stage moment conditions is:

$$E\left[\sum_{s=1}^{S}Pr_{z}\left(z_{b}=s|\left\{k_{bt},l_{bt},\phi_{bt}\right\}_{t=0}^{T};\boldsymbol{\Theta}\right)\left(\eta_{bt}^{s}-\frac{\sigma_{\eta,\zeta}^{s}}{\left(\sigma_{\zeta}^{s}\right)^{2}}\zeta_{bt}^{s}\right)\begin{pmatrix}\Delta_{\beta_{k}}\eta_{bt}\\ \Delta_{\beta_{l}}\eta_{bt}\\ 1\\ \omega_{bt-1}\end{pmatrix}\right]=0. \qquad (C.20)$$

C.3.3 Labor as a predetermined input

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

$$p_{\phi}^{s}(\phi_{bt}|k_{bt}, l_{bt}, \phi_{bt-1}, l_{bt-1}, k_{bt-1}; \boldsymbol{\beta}, \boldsymbol{\alpha}^{s}, \sigma_{\eta}^{s}) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi \left(\sigma_{\eta}^{s}\right)^{2}}} e^{-\frac{1}{2}\left(\frac{\eta_{bt}^{s}}{\sigma_{\eta}^{s}}\right)^{2}}, \tag{C.21}$$

which significantly simplifies the estimation procedure described above.

For comparison with the GMM estimation methodology specified above, we write down unconditional moment conditions related to the second-stage log-likelihood estimation. For a Cobb-Douglas production function with a linear Markov process for productivity, the set of four second-stage moment conditions is:

$$E\left[\sum_{s=1}^{S} Pr_{z}(z_{b} = s| \left\{k_{bt}, l_{bt}, \phi_{bt}\right\}_{t=0}^{T}; \boldsymbol{\Theta})\eta_{bt} \begin{pmatrix} \Delta_{\beta_{k}}\eta_{bt} \\ \Delta_{\beta_{l}}\eta_{bt} \\ 1 \\ \omega_{bt-1} \end{pmatrix}\right] = 0. \tag{C.22}$$

C.4 Comparison between estimation approaches

We compare in this section the impact on the estimated production function elasticities implied by the different approaches discussed above. We do so by estimating a Cobb-Douglas production function with a single-cluster linear Markov process for productivity with different moment conditions imposed (See Table C.1). Specifically, we differentiate between moment conditions imposed by GMM with predetermined labor (GMM Exo.) and ML. Additionaly, we differentiate between moment conditions imposed by GMM with non-predetermined labor (GMM Endo.), GMM where we use the control function approach to control for the endogeneity problem (GMM control function approach), and LIML. We expect to see differences in production function elasticities between the non-parametric GMM Endo. approach and the GMM control function approach if the parametric constraints imposed by the control function approach are restrictive. Similarly, if we observe differences between the GMM Exo. and ML, or the GMM control function approach and LIML, these differences can be ascribed to the differences in instrument specification for the moment conditions.

We apply our estimation procedure to Chilean data covering all manufacturing plants with more than 10 employees between 1979 and 1996, used by Gandhi et al. (2020) and sourced from Gandhi et al. (2020a). The results are displayed in Figures C.1 and C.2. We can observe that the main jump in the output elasticities can be observed from the assumption of whether labor is predetermined or not. Also, differences in instrument specification for the moment conditions has an influence on the resulting output elasticities, as can be deduced from comparing GMM exo. with ML, and GMM control function approach

Table C.1: Overview of unconditional moment conditions

Approach	Moment condition
GMM Exo.	$\mathrm{E}\left[\eta_{bt}\begin{pmatrix}k_{bt}\\ l_{bt}\\ 1\\ \omega_{bt-1}\end{pmatrix}\right]=0$
ML	$\mathrm{E}\left[\eta_{bt}\begin{pmatrix}\partial_{\beta_k}\eta_{bt}\\ \partial_{\beta_l}\eta_{bt}\\ 1\\ \omega_{bt-1}\end{pmatrix}\right]=0$
GMM Endo.	$\mathrm{E}\left[\eta_{bt}\begin{pmatrix}k_{bt}\\ l_{bt-1}\\ 1\\ \omega_{bt-1}\end{pmatrix}\right]=0$
GMM control function approach	$\mathrm{E}\left[\left(\eta_{bt}-\frac{\sigma_{\eta,\zeta}}{(\sigma_\zeta)^2}\zeta_{bt}\right)\begin{pmatrix}k_{bt}\\ l_{bt}\\ 1\\ \omega_{bt-1}\end{pmatrix}\right]=0$
LIML	$\mathrm{E}\left[\left(\eta_{bt}-\frac{\sigma_{\eta,\zeta}}{(\sigma_\zeta)^2}\zeta_{bt}\right)\begin{pmatrix}\partial_{\beta_k}\eta_{bt}\\ \partial_{\beta_l}\eta_{bt}\\ 1\\ \omega_{bt-1}\end{pmatrix}\right]=0$

Note: The presented moment conditions represent the just-identified unconditional moment conditions for a Cobb-Douglas production function with a sing-cluster linear Markov process for productivity.

with LIML. On the other hand, differences in output elasticities between the GMM Endo. approach and GMM control function approach are minimal, signaling that the functional form restrictions to resolve the endogeneity issue are not very restrictive in this setup.

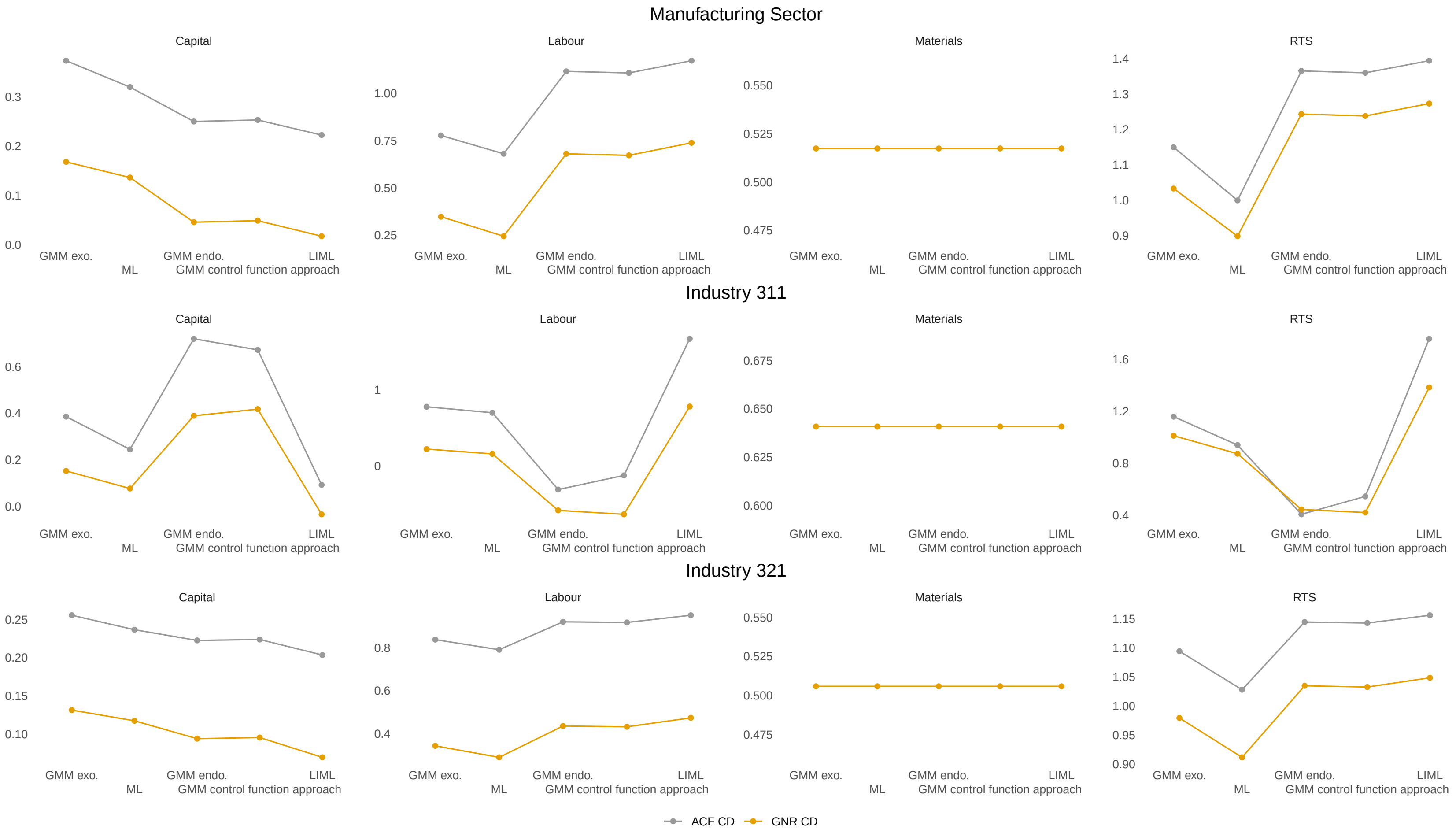

Figure C.1: Change in output elasticities, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas production function estimators with different moment conditions for the entire Manufacturing Sector and industries 311 and 321 of the Chilean economy.

Note: GMM Endo. and GMM Exo. refer to the Generalized Method of Moments estimation procedure with current labor and lagged labor as instruments, respectively. GMM control function approach refers to the Generalized Method of Moments estimation procedure where we model the endogeneity in the error term by incorporating a reduced-form specification for endogenous labor. ML and LIML refer to the Maximum Likelihood and Limited Information Maximum Likelihood estimation procedures, respectively.

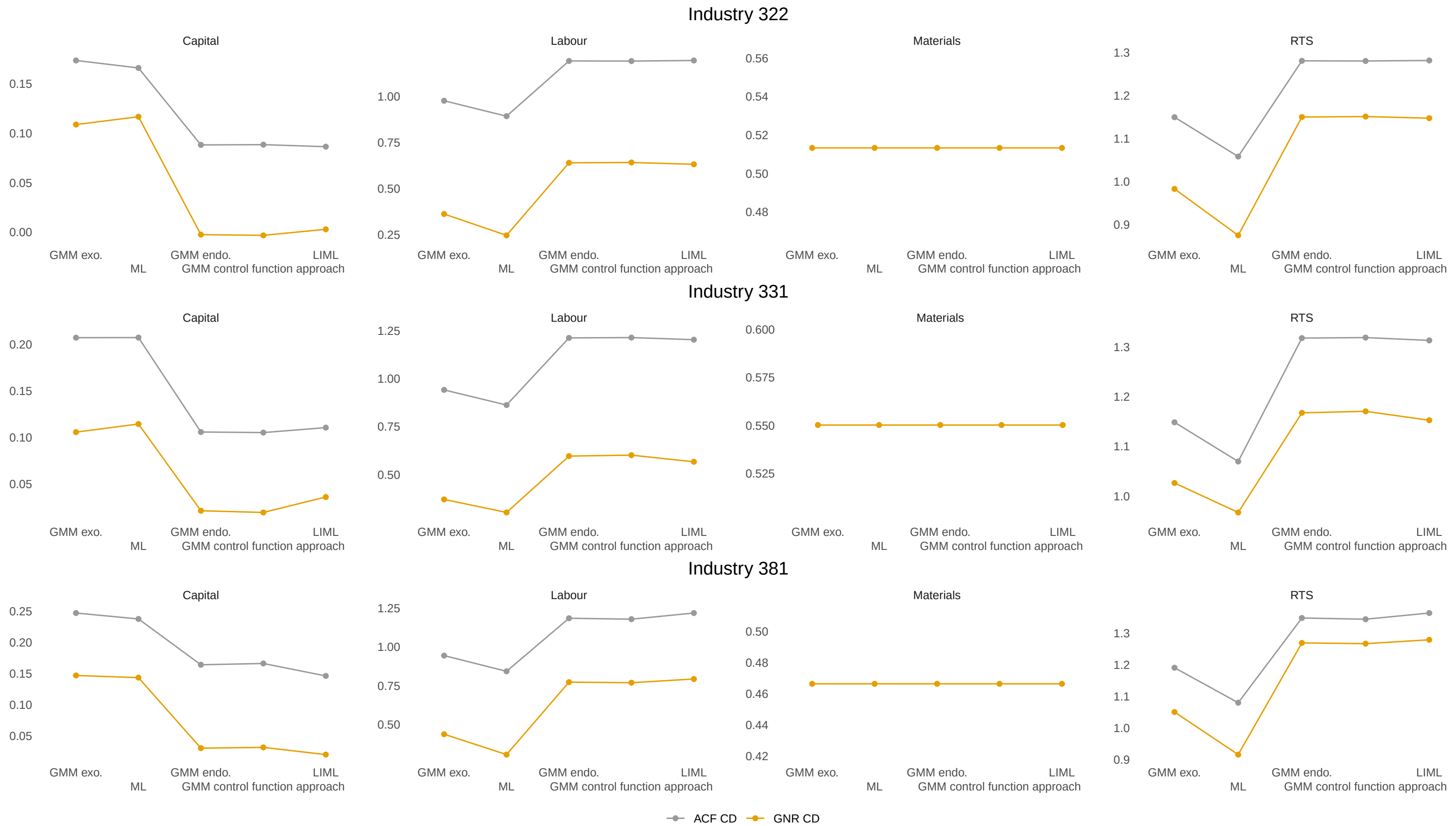

Figure C.2: Change in output elasticities, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas production function estimators with different moment conditions for the entire Manufacturing Sector and industries 322, 331, and 381 of the Chilean economy.
Note: GMM Endo. and GMM Exo. refer to the Generalized Method of Moments estimation procedure with current labor and lagged labor as instruments, respectively. GMM control function approach refers to the Generalized Method of Moments estimation procedure where we model the endogeneity in the error term by incorporating a reduced-form specification for endogenous labor. ML and LIML refer to the Maximum Likelihood and Limited Information Maximum Likelihood estimation procedures, respectively.

Appendix D Monte Carlo

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

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

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

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

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

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

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

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

$$\omega_{b,t-i} = \sum_{s=1}^{2} \mathbb{I}\left[z_b = s\right]\left[\alpha_0^s + \left(\alpha_1^s\right)^{1-i}\omega_{b,t-1} + \eta_{bt}^{c,A}\right];$$

$$\omega_{bt} = \mathbb{I}\left[z_b = s\right]\sum_{c=1}^{2}\left[\left(1-\left(\alpha_1^s\right)^i\right)\alpha_0^s + \left(\alpha_1^s\right)^i\omega_{b,t-i} + \eta_{bt}^{c,B}\right]. \tag{D.3}$$

Thus, when $i > 0$, firms have less than perfect information about ω_{bt} when choosing L_{bt}, and when i increases, this information decreases. Note that this specification is consistent with the finite mixture AR(1)-process specified in (D.2) since $\left(1-\left(\alpha_1^s\right)^i\right)\alpha_0^s + \left(\alpha_1^s\right)^i\alpha_0^s = \alpha_0^s$ 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:

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

where the variance of the normally distributed innovation η_{bt}^W $\left(\sigma^2_{\eta^W}\right)$ and the initial value $\ln\left(W_{b0}\right)$ are set in such a way that the standard deviation of $\ln\left(W_{bt}\right)$ is constant over time and equal to 0.1. Relative to a baseline in which all firms face the mean log wage in every period, this wage variation increases the within-firm, across-time, standard deviation of $\ln\left(L_{bt}\right)$ by about 10% (Ackerberg et al., 2015).

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

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

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

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

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

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

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

Appendix E Robustness

E.1 Estimation methodology and cluster selection

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

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

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

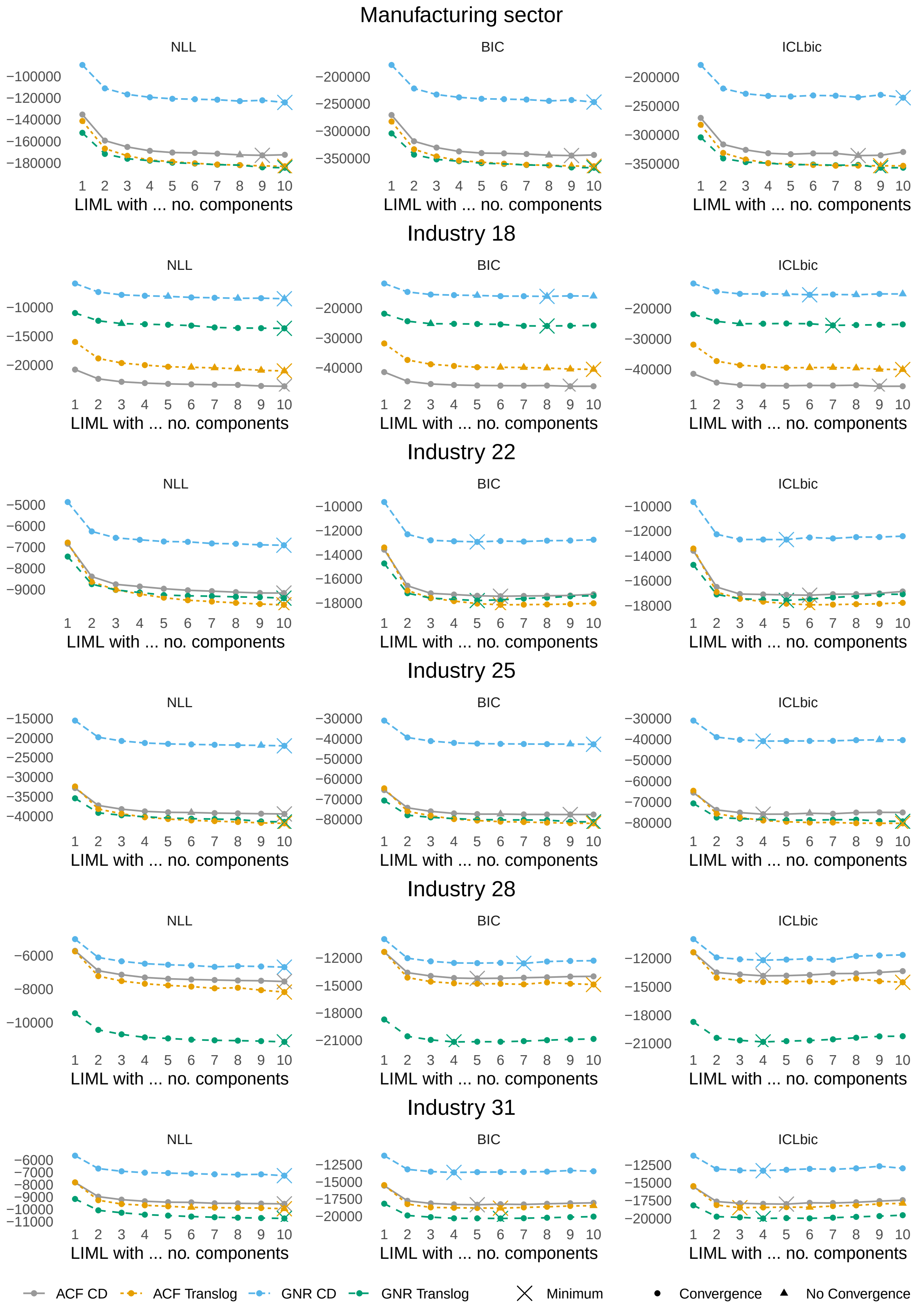

Figure E.1: Change in goodness-of-fit indicators of Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with endogenous labor in function of the number of clusters for the entire manufacturing sector and industries 18, 22, 25, 28, and 31 of the Belgian economy.
Note: NLL stands for negative log-likelihood, BIC for the Bayesian information criterion, and ICLbic for the integrated complete-data likelihood Bayesian information criterion. The times symbol indicates the optimal number of clusters defined by the minimum of the respective goodness-of-fit indicator. "No convergence" indicates non-convergence of the maximum likelihood estimation algorithm.

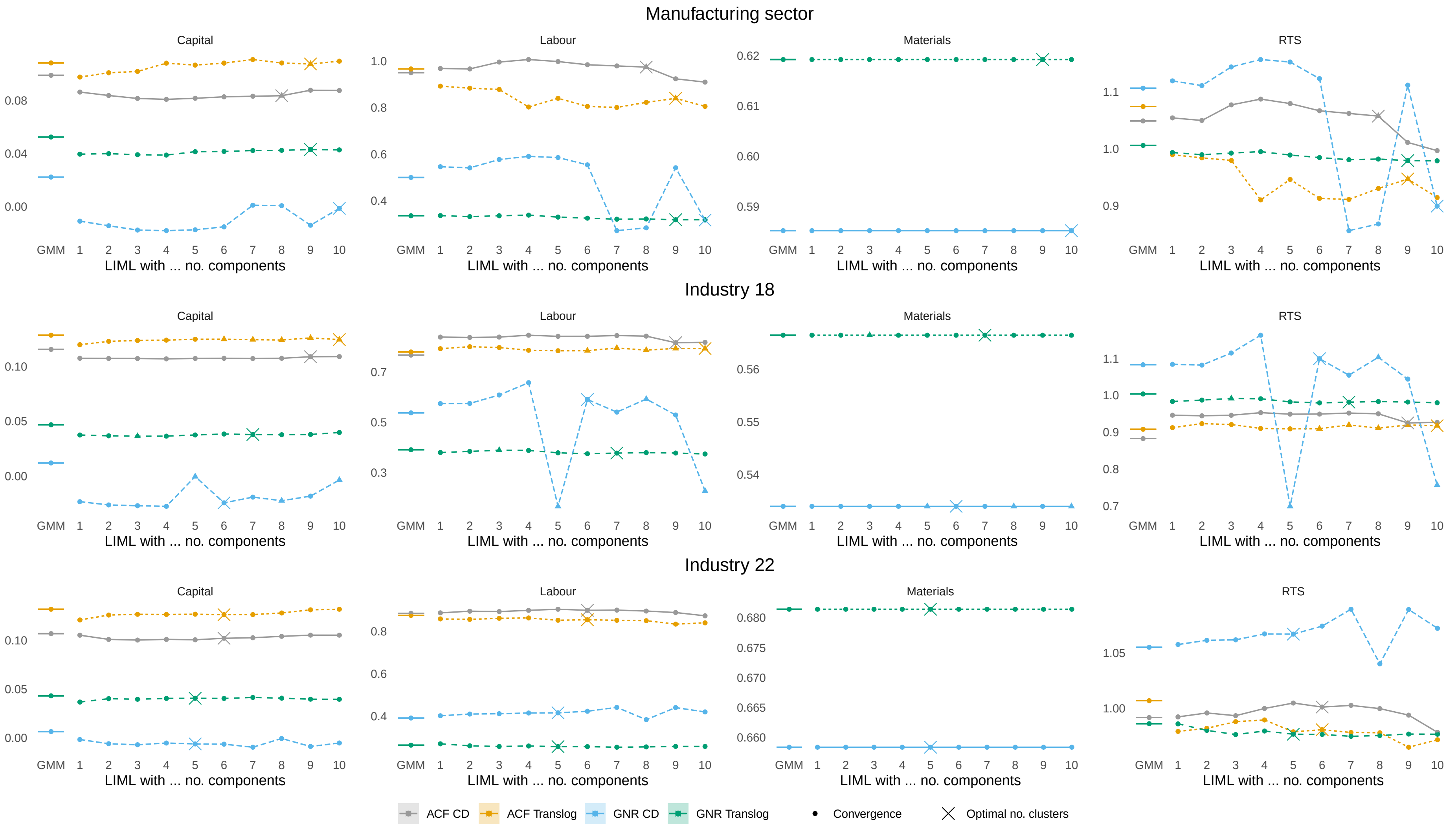

Figure E.2: Change in output elasticities, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with endogenous labor, in function of the number of clusters for the entire manufacturing sector and industries 18 and 22 of the Belgian economy.

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

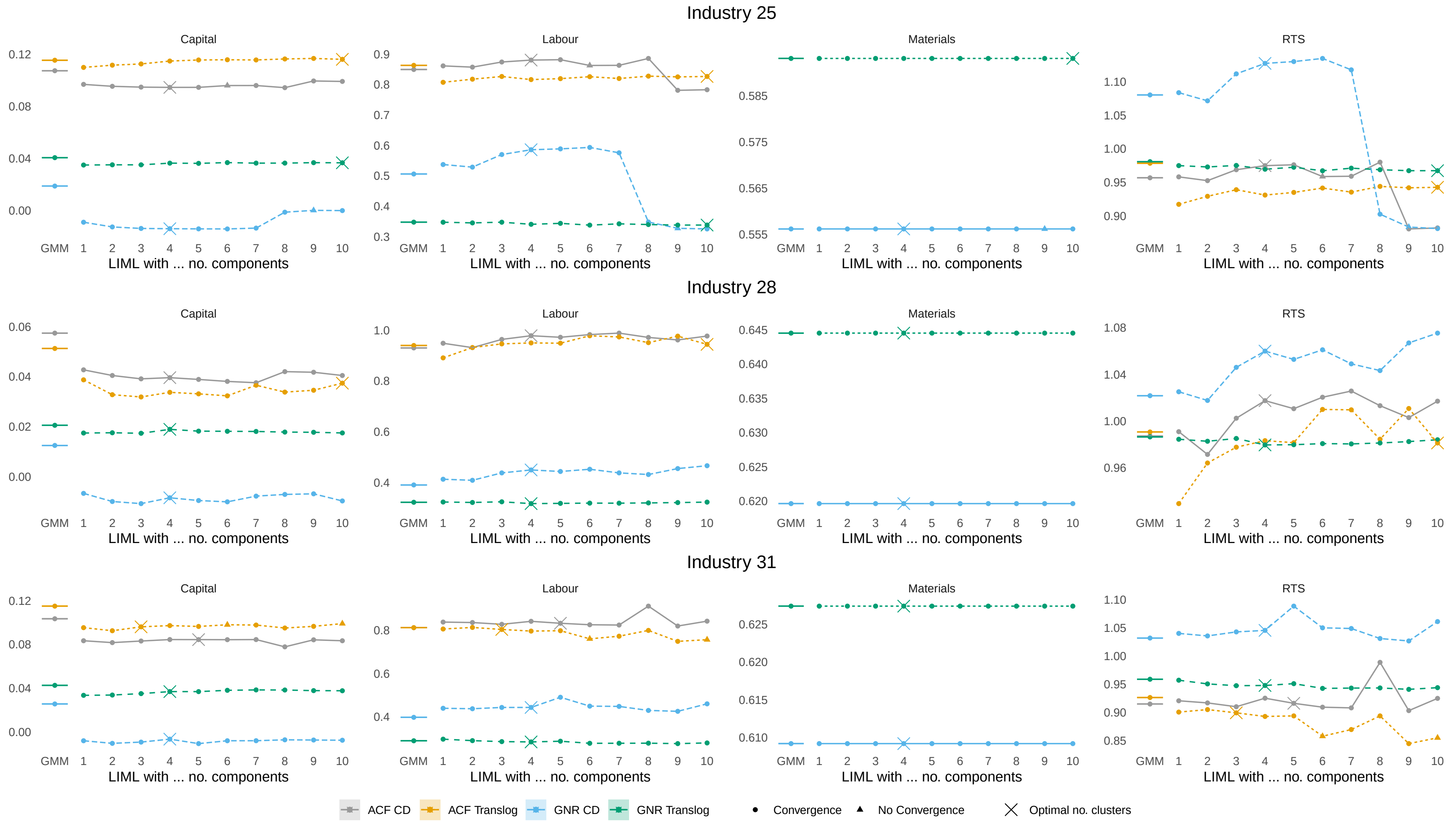

Figure E.3: Change in output elasticities, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with endogenous labor, in function of the number of clusters for industries 25, 28, and 31 of the Belgian economy.
Note: GMM and LIML refer to the generalized method of moments and limited information maximum likelihood as estimation procedure. The times symbol indicates the optimal number of clusters defined by the integrated complete-data likelihood Bayesian information criterion. "No convergence" indicates non-convergence of the maximum likelihood estimation algorithm.

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

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

Notes: a. The base specification refers to eq. (11), the augmented specification refers to eq. (21), and the specification without initial capital and labor refers to eq. (21) without initial capital and labor.
b. BIC stands for the Bayesian information criterion and ICLbic for the integrated complete-data likelihood Bayesian information criterion.

E.2 Chilean manufacturing sector

To evaluate the generalizability of the proposed productivity estimation methodology and the robustness of the reported results for the Belgian manufacturing sector, we apply our estimation procedure to Chilean data covering all manufacturing plants with more than 10 employees between 1979 and 1996, used by Gandhi et al. (2020) and sourced from Gandhi et al. (2020a). We follow Gandhi et al. (2020) in estimating the production function assuming labor is predetermined and, therefore, exogenous. In line with the main results, the goodness of fit statistics displayed in Figure E.4 provide evidence of heterogeneity in productivity. The fact that we find less heterogeneity in terms of the optimal number of clusters relative to the Belgian data can be ascribed to the data coverage. Whereas the Chilean data covers firms with more than 10 employees, the Belgian data covers all firms with more than one FTE employee. The production function estimates presented in Figures E.5 and E.6 are close to those obtained with current state-of-the-art estimation methodologies.

Next, we test the assumption that the initial conditions contain sufficient information to identify cluster affiliation for the Chilean case. If this assumption fails to hold, augmenting the base specification with additional, economically relevant firm-level characteristics is necessary to help improve the identification of cluster affiliation. To test this hypothesis, we augment equation (11) to the following multinomial logistic specification:

$$ln\frac{Pr_z(z_b = s|k_{b0}, l_{b0}, \omega_{b0}, \boldsymbol{e}_b; \boldsymbol{\gamma}^s)}{Pr_z(z_b = 1|k_{b0}, l_{b0}, \omega_{b0}, \boldsymbol{e}_b; \boldsymbol{\gamma}^1)} = \gamma_0^s + \gamma_1^s k_{b0} + \gamma_2^s l_{b0} + \gamma_3^s \omega_{b0} + \gamma_4^s EXP_b + \gamma_5^s IMP_b + \gamma_6^s Hiwag_b + \gamma_7^s Adv_b, \quad \forall s = 2, \ldots, S \tag{E.1}$$

where cluster probabilities are specified conditional on initial capital, labor, and productivity as well as additional firm characteristics represented in the vector $\boldsymbol{e}_b = \{EXP_b, IMP_b, Hiwag_b, Adv_b\}$, such as indicators of export (EXP_b) and import activity (IMP_b), of firms paying a higher wage than the industry median ($hiwag_b$), and of firms reporting advertisement expenditures (Adv_b) over the sample period.[2] Furthermore, we specify a version of equation (E.1) without initial capital, labor, and productivity. If the considered firm-level characteristics contain sufficient information to group firms into clusters, we expect this specification to perform as well as our base specification.

We rely on the two augmented specifications discussed above to re-estimate the production function. The resulting log-likelihood, BIC, and ICLbic are reported in Table E.2. In line with the main results we find that, with the exception of industry 311, the increase in log-likelihood obtained by the augmented specification is insufficient to warrant

[2] We set the firm-level indicators equal to zero when they report no activity in this area over the entire sample period or report activity at least one point in time during the sample period. Firms reporting no activity for all indicators are chosen as the reference group.

the increase in the number of parameters, as indicated by the smaller BIC and ICLbic indicators in absolute value relative to the base specification. Additionally, we reaffirm the importance of the initial conditions in identifying cluster affiliation. Comparing the model fit for the augmented specification without initial conditions to the base specification, the ICLbic and BIC indicate that the latter performs best. Therefore, even when firm-level information regarding the internationalization status of a firm, its advertising expenditure, and the relative height of its wages is available, a significant share of the heterogeneity in productivity remains latent and cannot be accounted for using existing methods in the literature.

Lastly, we calculate the decomposition of aggregate productivity for different Chilean productivity indices obtained from different estimation methodologies and different specifications of heterogeneity in productivity.[3] Specifically, we estimate productivity using the GMM and ML identification strategies with (i) a *base* specification: $\omega_{bt} = \alpha_0 + \alpha_1 \omega_{bt-1} + \eta_{bt}$, (ii) a *deterministic* control for exporter status, $\omega_{bt} = \alpha_0 + \alpha_1 \omega_{bt-1} + \alpha_2 EXP_b + \alpha_3 \omega_{bt-1} EXP_b + \eta_{bt}$, and (iii) a more *exhaustive* set of controls for heterogeneity in productivity:

$$\begin{aligned} \omega_{bt} = \alpha_0 + \alpha_1 \omega_{bt-1} &+ \alpha_4 EXP_b + \alpha_5 \omega_{bt-1} EXP_b \\ &+ \alpha_6 IMP_b + \alpha_7 \omega_{bt-1} IMP_b \\ &+ \alpha_8 Hiwag_b + \alpha_9 \omega_{bt-1} Hiwag_b \\ &+ \alpha_{10} Adv_b + \alpha_{11} \omega_{bt-1} Adv_b + \eta_{bt}. \end{aligned} \tag{E.2}$$

Similarly, we obtain productivity from the finite mixture ML identification strategy with the optimal number of four clusters and (i) the *base* specification for cluster affiliation (11), (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 (E.1).

Figure E.7 displays the evolution of the obtained aggregate productivities and their decomposition across estimation methodologies and specifications for Industry 311 with the optimal number of four clusters in Chile. In line with the main results, we observe that the evolution of aggregate productivity over time is very similar across estimation methodologies and specifications. Moreover, the dependence of the export premium on the specification for the GMM and ML estimation methodology is more pronounced in Chile than in Belgium. For instance, the export premium for Industry 311 —the difference between the average productivity of exporters (dashed line) and non-exporters (continuous line) in the second column of Figure E.7— evolves from 12.18%, for the base

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

specification, to 22.46% for the deterministic and 49.57% for the exhaustive specification of heterogeneity for the GMM methodology (see Table E.3). In comparison, the export premium is approximately 76% for all three specifications of the finite mixture ML.

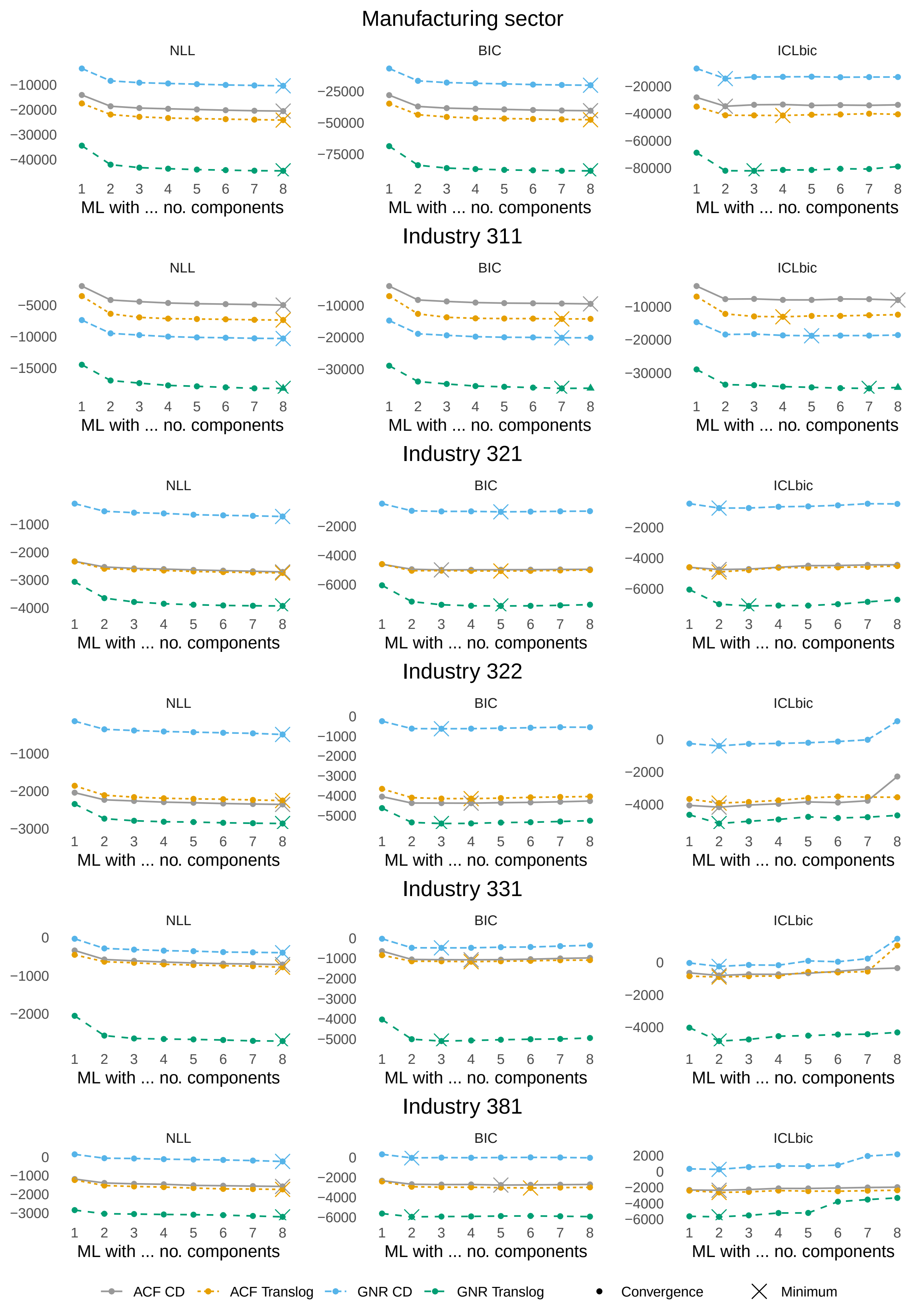

Figure E.4: Change of goodness-of-fit indicators of Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with exogenous labor, in function of the number of clusters for the entire manufacturing sector and industries 311, 321, 322, 331, and 381 of the Chilean economy.

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

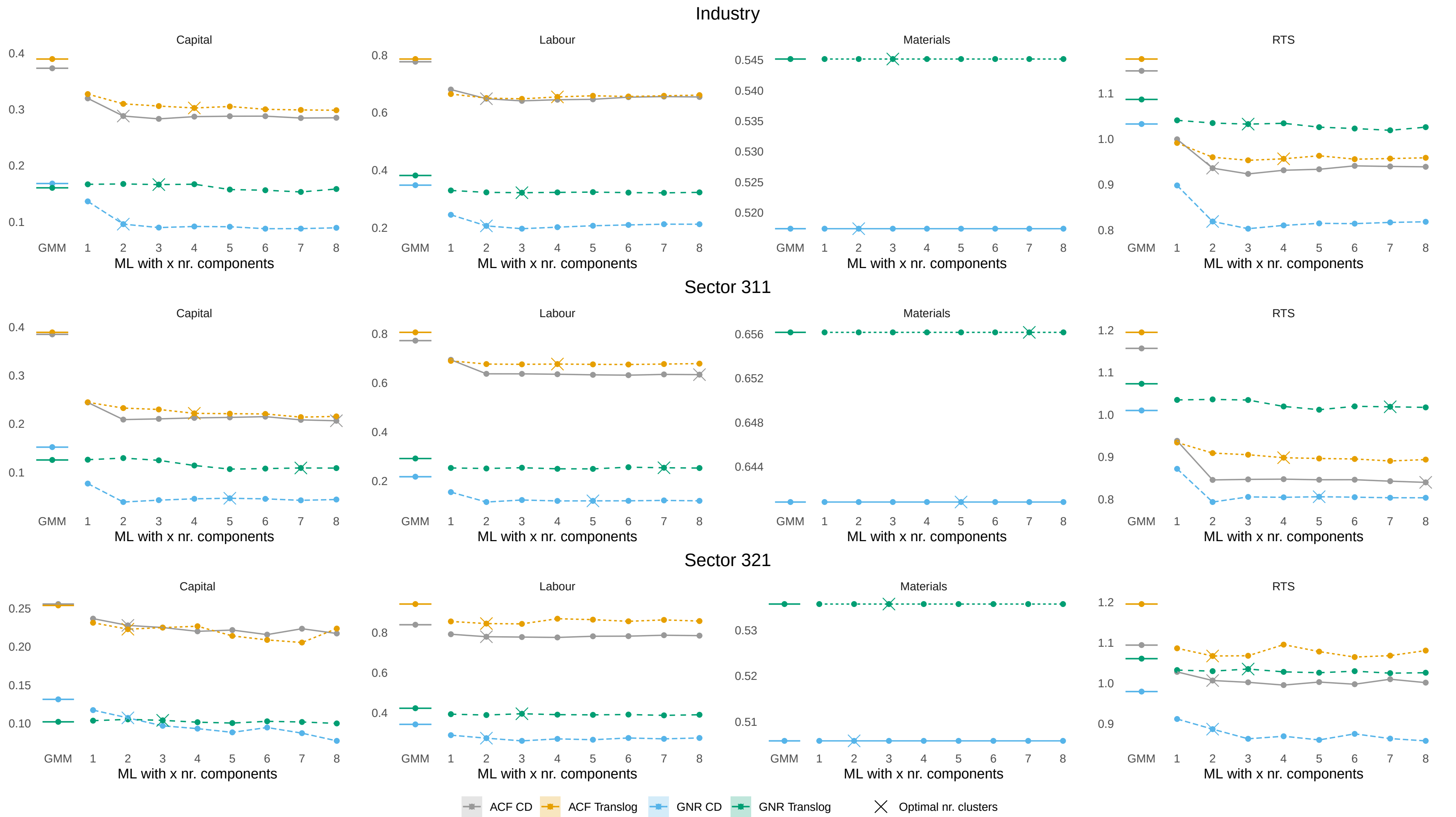

Figure E.5: Change in output elasticities, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with exogenous labor, in function of the number of clusters for the entire manufacturing sector and industries 311, 321 of the Chilean economy.

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

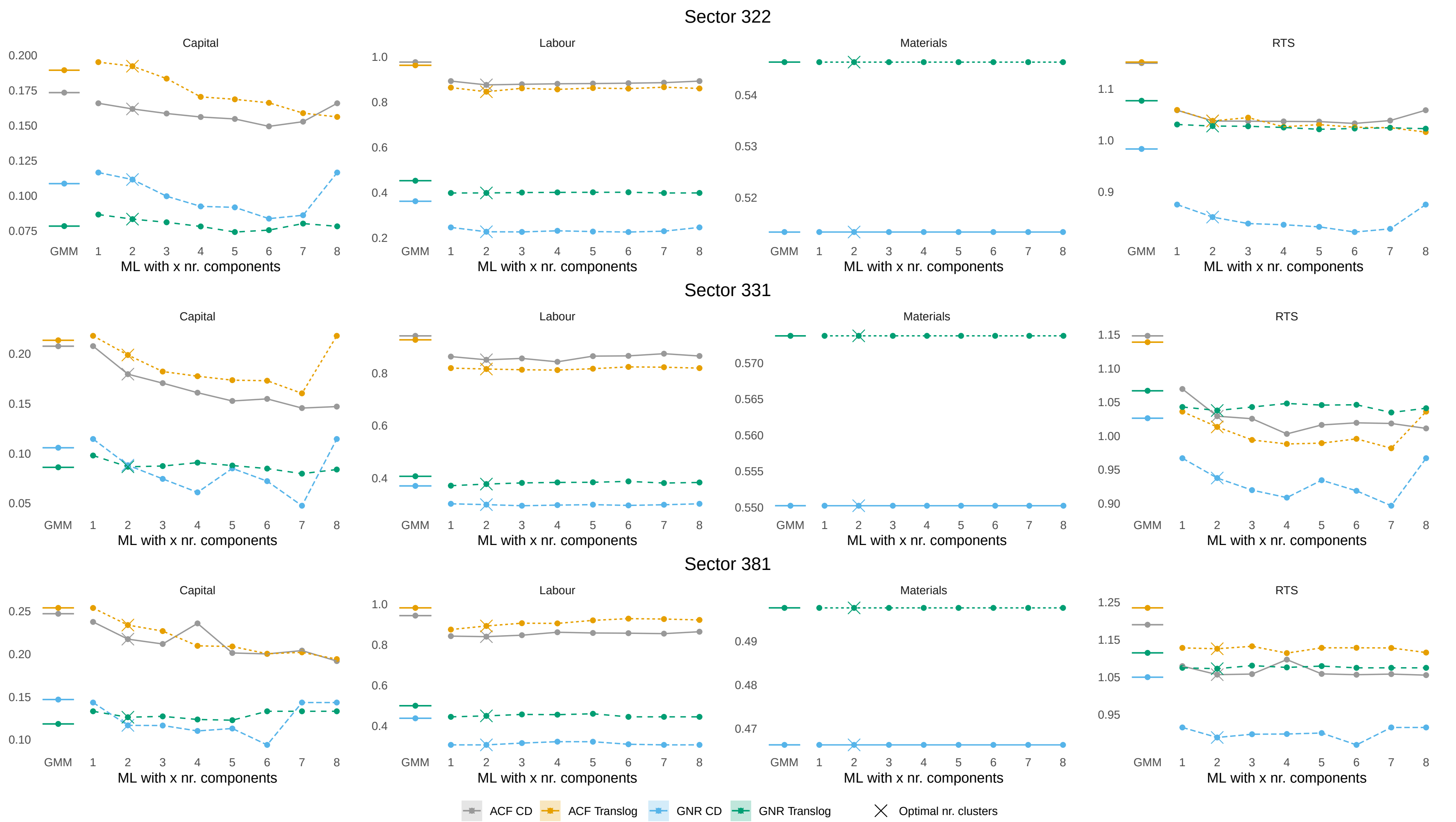

Figure E.6: Change in output elasticities, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with exogenous labor, in function of the number of clusters for industries 322, 331, and 381 of the Chilean economy.
Note: GMM and ML refer to the generalized method of moments and maximum likelihood estimation procedures. The times symbol indicates the optimal number of clusters as defined by the integrated complete-data likelihood Bayesian information criterion.

Table E.2: Goodness-of-fit indicators for estimation with varying concomitant specifications.

Specification	Log-likelihood	BIC	ICLbic
Industry 311			
Base specification	7,155.82	-14,031.99	-13,121.45
Additional concomitants	7,260.46	-14,125.55	-13,293.26
Without initial capital and labor	7,134.40	-13,767.35	-12,277.98
Industry 321			
Base specification	2,597.86	-5,069.94	-4,921.05
Additional concomitants	2,600.65	-5,041.97	-4,887.88
Without initial capital and labor	2,638.43	-5,042.05	-4,891.89
Industry 322			
Base specification	2,113.01	-4,103.07	-3,909.68
Additional concomitants	2,115.32	-4,074.91	-3,878.02
Without initial capital and labor	2,129.69	-4,029.88	-3,828.29
Industry 331			
Base specification	636.73	-1,150.08	-900.91
Additional concomitants	639.29	-1,122.31	-878.73
Without initial capital and labor	657.56	-1,084.82	-878.97
Industry 381			
Base specification	1,537.71	-2,949.10	-2,647.88
Additional concomitants	1,539.18	-2,918.37	-2,619.81
Without initial capital and labor	1,539.57	-2,843.34	-2,527.09

Notes: The base specification refers to equation (11), the augmented specification refers to equation (E.1), and the specification without initial capital and labor refers to equation (E.1) without initial capital and labor. BIC stands for Bayesian Information Criterion and ICLbic for Integrated Complete-data Likelihood Bayesian Information Criterion. Estimates are obtained from a Value-added Translog production function with exogenous labor.

Figure E.7: Evolution of aggregate productivity and its decomposition for exporting- and non-exporting firms for Industry 311 in Chile.

Notes: GMM, ML, and 4-cluster finite mixture ML refer to the productivity estimation methodologies, while Base, Deterministic, and Exhaustive refer to the specification of heterogeneity within these methodologies, i.e., see equations (E.2) and (E.1).

Table E.3: Average Export Premia across productivity estimation methodologies and specifications

Methodology	Specification	Industry 311	Industry 321	Industry 322	Industry 331	Industry 381
GMM	Base	0.1218	0.0588	0.1542	0.0341	0.0309
		(0.0097)	(0.0070)	(0.0091)	(0.0115)	(0.0054)
GMM	Deterministic	0.2246	0.0866	0.1727	0.0701	0.0372
		(0.0091)	(0.0073)	(0.0090)	(0.0119)	(0.0054)
GMM	Exhaustive	0.4957	0.1067	0.1982	0.1018	0.0671
		(0.0085)	(0.0076)	(0.0088)	(0.0123)	(0.0053)
ML	Base	0.6664	0.1451	0.2421	0.0840	0.1008
		(0.0095)	(0.0074)	(0.0098)	(0.0118)	(0.0046)
ML	Deterministic	0.7552	0.1743	0.2506	0.1273	0.1144
		(0.0092)	(0.0078)	(0.0097)	(0.0122)	(0.0045)
ML	Exhaustive	0.7484	0.1845	0.2605	0.1553	0.1434
		(0.0090)	(0.0079)	(0.0097)	(0.0125)	(0.0043)
Finite Mixture ML	Base	0.7630	0.1733	0.2506	0.1382	0.1184
		(0.0095)	(0.0078)	(0.0102)	(0.0124)	(0.0045)
Finite Mixture ML	Deterministic	0.7654	0.1734	0.2505	0.1383	0.1184
		(0.0095)	(0.0078)	(0.0102)	(0.0124)	(0.0045)
Finite Mixture ML	Exhaustive	0.7675	0.1728	0.2506	0.1369	0.1180
		(0.0095)	(0.0078)	(0.0102)	(0.0124)	(0.0046)

Notes: Export premia obtained from a log-linear regression with year dummies t, $\omega_{bt} = \alpha Exp_b + t + \epsilon_{bt}$, where productivity is obtained from a GMM, ML and Finite Mixture ML estimation methodology with a Base, Deterministic, or Exhaustive specification of heterogeneity within these methodologies, i.e. see (E.2) and (E.1). Standard Errors between brackets obtained from the OLS regression.

Appendix F Robustness of Production Function Coefficients

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

First, the identification strategy in this paper relies on random cluster affiliation, in contrast to the deterministic cluster affiliation currently used in the literature. Despite having demonstrated the adequacy of the random cluster affiliation identification strategy in the Monte Carlo exercise (see Section 3.4), we additionally evaluate the robustness of production function coefficients in our Belgian firm-level data using deterministic cluster identification strategies. To this end, we estimate separate production functions for 5 NACE Rev.2 industries, which are: Printing and reproduction of recorded media (18); Manufacture of rubber and plastic products (22); Manufacture of fabricated metal products, except machinery and equipment (25); Manufacture of machinery and equipment n.e.c. (28); and Manufacture of furniture (31), and an aggregate production function for the entire manufacturing sector. We parameterize the production function $f(\cdot;\boldsymbol{\beta})$ assuming both a gross-output (Gandhi et al., 2020) and value-added (Ackerberg et al., 2015) production function under both a Cobb-Douglas and Translog specification. These production functions are estimated using a GMM estimation approach with either a simple linear Markov process specification:

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

or a deterministic Markov specification:

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

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

The results of this exercise, presented in Figures F.1 and F.2, confirm the evidence of robust production function coefficients in our dataset. We observe no significant deviations between the output elasticities obtained from a linear Markov process (None) and those obtained from a Markov process allowing for heterogeneity depending on whether the firm is respectively an exporter, importer, engaged in FDI, or all three simultaneously

(Export, Import, FDI, All).

Second, this paper relies on the commonly used scalar unobservability assumption, stating that materials are a flexible factor input that is decided upon simultaneously at time t without affecting future profits ($m_{bt} = h(\omega_{bt}, k_{bt}, l_{bt})$). Under the scalar unobservability assumption, the change in future productivity (and cluster affiliation) does not affect the choice of material inputs (Ackerberg, 2021). It could be, however, that a firm's cluster affiliation affects its input demand. For instance, a firm's export status has been argued to lead to differences in optimal input demand across firms (De Loecker and Warzynski, 2012). Suppose this export status is a determinant of cluster affiliation. In that case, the cluster affiliation might then also affect optimal input demand, such that $m_{bt} = h(\omega_{bt}, k_{bt}, l_{bt}, z_b^s)$. This would be in line with Kasahara et al. (2023) assuming the material input demand depends on cluster affiliation. Shenoy (2020) provides a formal framework to evaluate the adequacy of the scalar unobservability assumption. The author demonstrates that failing to account for relevant variables affecting input demand is equivalent to introducing a non-classical measurement error in the first stage of the production function estimation procedure. If this is the case, our first-stage estimation procedure might be misspecified, with unobserved heterogeneity largely being captured by the first-stage residual ε_{bt}. This could explain the robustness of our second-stage production function estimation results to the homogeneity assumption of productivity growth. However, it is unclear why one would argue that the FOC for the perfectly flexible input may be cluster-dependent while FOCs of the non-flexible inputs are not. Cluster-dependent FOCs for all inputs imply a cluster-specific production function specification, which falls outside the scope of this paper. In the concluding Section 6, we discuss the possibilities that the methodology proposed in this paper opens for future research, including cluster-dependent production function specifications.

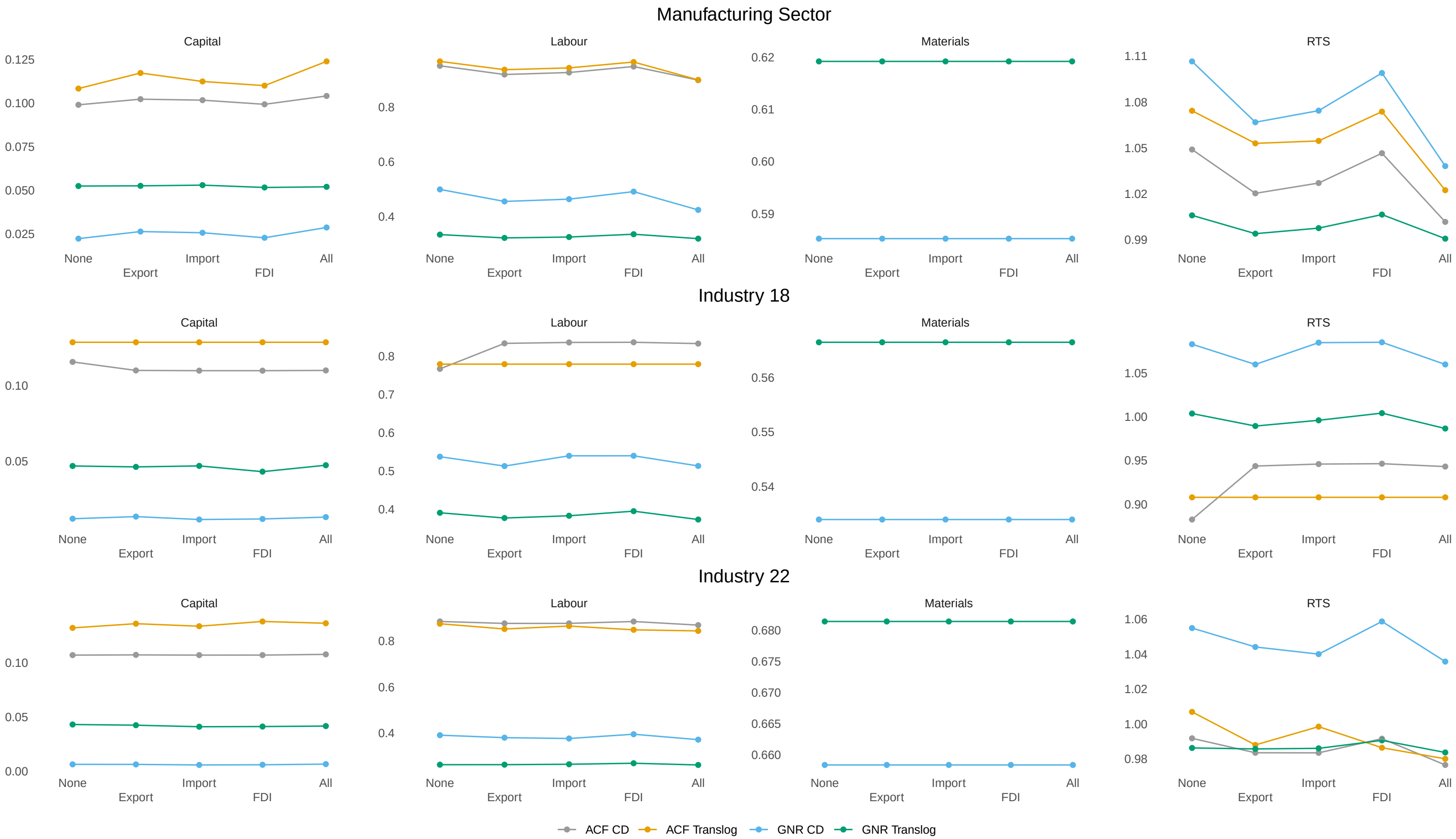

Figure F.1: Change in output elasticities, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with endogenous labor, in function of the Markov specification for the complete Industry and industries 18 and 22 of the Belgian economy.

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

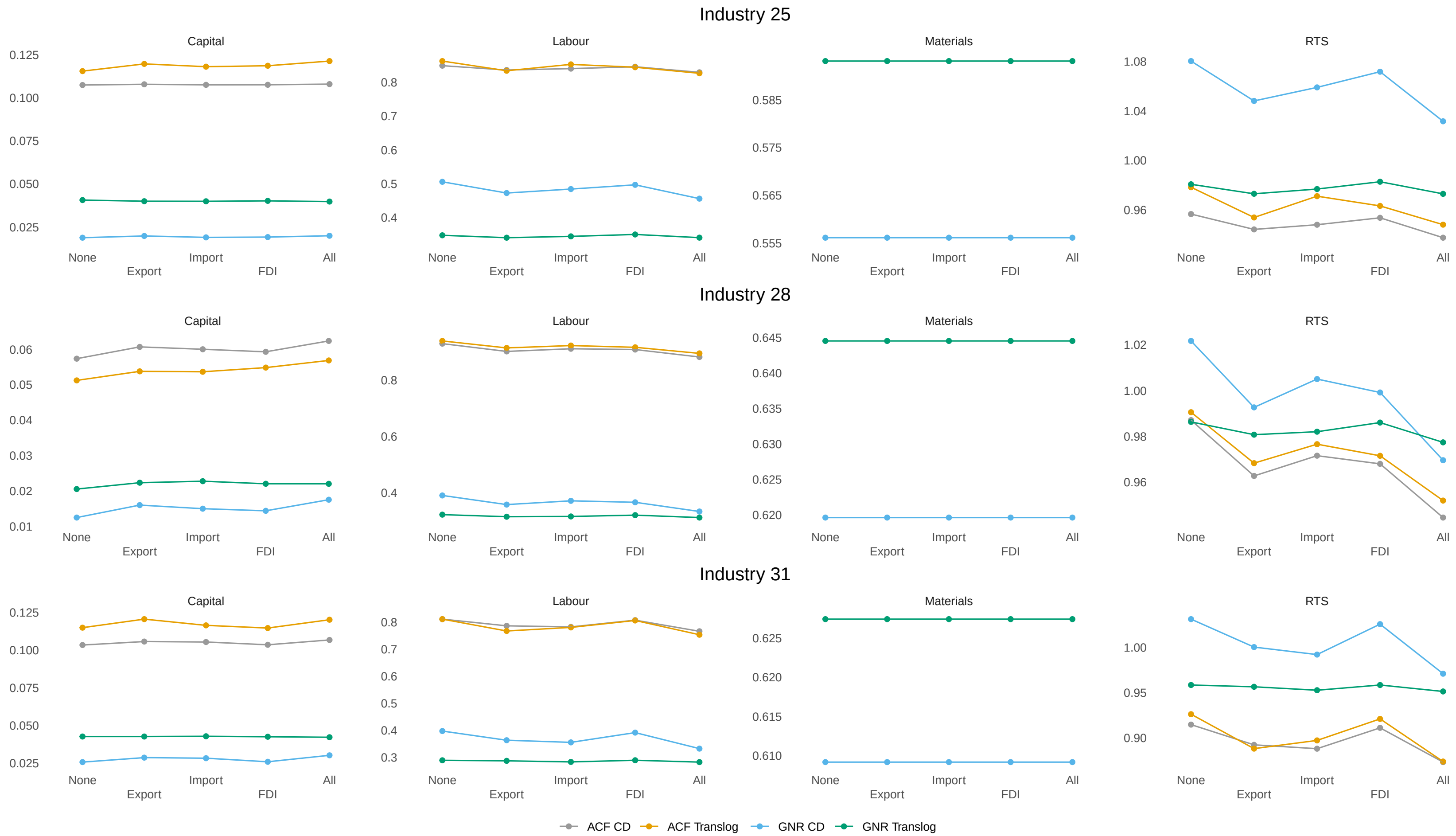

Figure F.2: Change in output elasticities, based on Ackerberg et al. (2015) and Gandhi et al. (2020) Cobb-Douglas and Translog production function estimators with endogenous labor, in function of the Markov specification for the industries 25 and 28, and 31 of the Belgian economy.

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

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 (2023, May). Identification and Estimation of Production Function with Unobserved Heterogeneity.

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.