WIDER TECHNICAL NOTE | 20/2021

Southern Africa – Towards Inclusive Economic Development (SA-TIED)

Total factor productivity in South African manufacturing firms 2010–17

CIT-IRP5 panel v4.0

C. Friedrich Kreuser[1] and Daniel Brink[2]

October 2021

Abstract: We update Kreuser and Newman's (2018) total factor productivity estimates for the South African manufacturing sector using administrative data from 2009–17. We use standard implementations of the Ackerberg et al. (2015) and Wooldridge (2009) productivity estimators and introduce an implementation of the former that converges to elasticities inside the unit interval more consistently. Limited productivity growth is found for the period 2010–17, with the majority of variance in sectoral productivity attributable to variance in allocative efficiency. We find that industries that have higher share sales allocated to more productive firms are generally more capital intensive and have significantly higher capital elasticities of output while only having slightly higher labour elasticities of output.

Key words: total factor productivity, tax administrative data

JEL classification: C61, D22, D24

Acknowledgements: We would like to thank the following members of the UNU-WIDER and South African National Treasury Team for their technical assistance and support: Amina Ebrahim, Marlies Piek, Michael Kilumelume, Michelle Pleace, Mlungisi Ndlovu, and the South African National Treasury. We are also grateful to Carol Newman for detailed comments on the paper. We would like to thank Rulof Burger for comments on the optimization routines. C. Friedrich Kreuser gratefully acknowledges support from UNU-WIDER for this project.

Supplementary material: The STATA.do files used to prepare this note are available on the technical note's webpage (https://www.wider.unu.edu/node/238197).

Related publications:

C. Kreuser, F. and C. Newman (2018). 'Total Factor Productivity in South African Manufacturing Firms'. *South African Journal of Economics*, 86(S1): 40–78. https://doi.org/10.1111/saje.12179

1 PhD candidate, Department of Economics, Trinity College Dublin, Ireland, corresponding author: kreuserc@tcd.ie;
2 Research assistant, UNU-WIDER and the National Treasury of South Africa, Pretoria, South Africa

This technical note has been prepared within the UNU-WIDER project Southern Africa – Towards Inclusive Economic Development (SA-TIED).

Information and requests: publications@wider.unu.edu

https://doi.org/10.35188/UNU-WIDER/WTN/2021-20

United Nations University World Institute for Development Economics Research

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

The United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland, Sweden, and the United Kingdom as well as earmarked contributions for specific projects from a variety of donors.

1 Introduction

This paper uses South African tax administrative data to update the total factor productivity (TFP) estimates for the country's manufacturing sector by Kreuser and Newman (2018) (KN). We expand the sample period to 2009–17 from KN's initial period of 2009–13. This study benefits from industry classification improvements made by Budlender and Ebrahim (2020), improved identification of employees by Kerr (2020), and general updates to the data made by Ebrahim et al. (2021).

We expand KN's approach by using both the Wooldridge (2009) and the Ackerberg et al. (2015) (ACF) approaches. We present a new optimization routine for the ACF estimator to overcome the convergence issues common in standard implementations, without resorting to increased data requirements as in Kim et al. (2019). Our implementation yields coefficients inside the unit interval for both capital and labour elasticities more consistently than other ACF implementations and appears to be more robust to changes in initial values and sample restrictions. Our procedure provides the coefficients for the temporal relationship of the productivity residuals implied by the ACF procedure as well as the minima found by the optimizer. The latter is reported to enable at least some goodness of fit evaluation of the estimates. Finally, we provide new capital stock measures for the firm based on the perpetual inventory approach with a series of imputations, adjustments, and alterations. We show that constructing a capital stock series using a straight line 10 per cent depreciation yields generally more consistent results with capital stock aggregates in line with those reported by external sources (StatsSA 2010–2018a). The code for our algorithm, data cleaning, and construction code is available with this paper.

We find weak productivity growth trends consistent with those found by KN. We expand KN's approach by decomposing productivity into allocative and technical efficiency and show that the majority of growth can be attributed to the former. We find evidence that the food, paper, and chemicals and pharmaceutical sectors have the highest share of output attributed to high productivity firms, meaning that they are the most allocatively efficient industries in our sample. We find further evidence of poor allocative efficiency in the fabricated metals, wood and cork, electrical equipment, and furniture sectors. The former three industries have been investigated by competition authorities.[1] Finally, we find evidence that industries with the highest allocative efficiency generally have higher capital elasticities of output, and that these industries have higher aggregate capital-to-labour and output-to-capital ratios.

This paper is organized as follows: Section 2 provides a detailed overview of the ACF estimator and discusses in detail the practical steps required in estimation as well as our preferred implementation. Section 3 discusses the construction of productivity measures used, and Section 4 discusses the data used. Section 5 discusses our results and Section 6 concludes.

2 Identification strategy

In this section, we outline the identification strategy by first discussing the functional form of interest and the shock processes. We then discuss the identification assumptions in the Olley and Pakes (1996) (OP), Levinsohn and Petrin (2003) (LP), and ACF approaches, after which we discuss the moment conditions implied by the latter. While we use the Wooldridge (2009) estimator, the reader is referred to KN for the discussion using the CIT-IRP5 data. This section draws heavily on ACF and Collard-Wexler and De Loecker (2020) in both content and structure.

[1] We discuss these industries in Section 5.3.

Equation (1) shows the gross output and Leontief type production functions in levels. In the equation, $Q_{i,t}$ is output, $K_{i,t}$ is fixed capital stock used in production, $L_{i,t}$ is labour input, $\Omega_{i,t}$ is Hicks-neutral productivity, $M_{i,t}$ is the materials input, and $\varepsilon_{i,t}$ is some i.i.d (independent and identically distributed) shock that affects the firm's output in the present period and is uncorrelated to all other past, present, and future values on the right-hand-side of the equation.

$$Q_{i,t} = \begin{cases} M_{i,t}^{\beta_m} L_{i,t}^{\beta_l} K_{i,t}^{\beta_k} \Omega_{i,t} Exp(\varepsilon_{i,t}) & (Gross-Output) \\ min(\beta_m M_{i,t}, L_{i,t}^{\beta_l} K_{i,t}^{\beta_k} \Omega_{i,t} Exp(\varepsilon_{i,t})) & (Leontief) \end{cases} \tag{1}$$

We can write the Hicks-neutral component of firm productivity as in Equation (2), where $\omega_{i,t}$ is the natural logarithm of firm productivity as a separable function of the firm's known constant productivity, β_i, and productivity shock process, $\eta_{i,t}$.

$$\Omega_{i,t} = Exp(\omega_{i,t}) = Exp(\beta_i + \eta_{i,t}) \tag{2}$$

We will treat the shock process for $\omega_{i,t}$ as the shock process for $\eta_{i,t}$. We assume that the the firm's fixed-effect information is known to the firm and thus used in its optimization problem, but that this fixed effect is orthogonal to the shock process. That is, we assume that $\eta_{i,t}$ follows a process such that $E[\eta_{i,t}|I_{-\beta_i,i,t}] = E[\eta_{i,t}|I_{i,t}]$, so that the constant term does not contribute any information to the firm about $\eta_{i,t}$'s value. We assume that $\eta_{i,t}$ follows the autoregressive of order 1 (AR(1)) process as in Equation (4) such that $E[\eta_{i,t}|\eta_{i,t-1},\beta_i] = E[\eta_{i,t}|\eta_{i,t-1}]$. The remainder of this paper will discuss the shock process in terms of $\omega_{i,t}$, to remain consistent with the literature (Ackerberg et al. 2015; Collard-Wexler and De Loecker 2020).

$$\omega_{i,t} = \beta_i + \eta_{i,t} \tag{3}$$

$$\eta_{i,t} = \rho\eta_{i,t-1} + \xi_{i,t} \tag{4}$$

Keeping the suppression of β_i in mind, the logarithmic version of the production function can be written as Equation (5), where $q_{i,t} = ln(Q_{i,t})$, $k_{i,t} = ln(K_{i,t})$, and $l_{i,t} = ln(L_{i,t})$.[2]

$$q_{i,t} = \beta_0 + \beta_l l_{i,t} + \beta_k k_{i,t} + \omega_{i,t} + \varepsilon_{i,t} \tag{5}$$

2.1 Assumptions

In this section we outline the assumptions made in the identification of parameters in the standard approaches and highlight the data requirements in the standard (ACF) approach. The data-generating process (DGP) considered by ACF allows for identification of production function parameters in perfectly competitive output markets and competitive input markets with common factor prices.[3]

Assumption 1: information set

For each period t the firm solves its objective function according to its information set at time t. The information set at time t, $I_{i,t}$, includes all current and past productivity shocks but not future shocks. Thus, firm can only form expectations of $\{\omega_{i,\tau}\}_{\tau=t+1}^{\infty}$, but has perfect information on $\{\omega_{i,\tau}\}_{\tau=0}^{t}$. The firm does not observe the transitory shock $\varepsilon_{i,t}$ until after all decisions have been made. The transitory shock has $E[\varepsilon_{i,t}] = 0$.

[2] Note here that β_0 is $E[q_{i,t}|l_{i,t} = 0, k_{i,t=0}]$. Where fixed effects exist in $\omega_{i,t}$, we can write the version of ω in (5) as $\omega_{i,t} = \beta_i - \beta_0 + \eta_{i,t}$.

[3] Collard-Wexler and De Loecker (2020) uses this approach to focus on measurement error in capital.

Assumption 2: shock process

Productivity shocks evolve according to a first order Markov process as in Equation (6). The general distribution of the process $p(.)$ is known to firms and is stochastically increasing in $\omega_{i,t}$. We treat this process as a simple AR(1) for exposition as in Equation (4), above, where $\xi_{i,t}$ is an unknown shock with $E[\xi_{i,t}|I_{i,t}] = 0$. Firms only observe $\omega_{i,t}$ at time t, after all decisions $t-b$ with $0 < b < \infty$ have already been made.

$$p(\omega_{i,t+1}|I_{i,t}) = p(\omega_{i,t+1}|\omega_{i,t}) \tag{6}$$

Assumption 3: timing of inputs

The firm's problem in the OP, LP, and ACF literature generally sets the firm's investment problem as being made at $t-1$, so that capital stock is always treated as a state variable in period t. Capital stock evolves according to the equation of motion in Equation (7), where $\varkappa$ is the solution to the firm's optimization problem.

$$k_{i,t} = \varkappa(k_{i,t-1}, i_{i,t-1}) \tag{7}$$

In the OP and LP literature, labour is a non-dynamic input, meaning that it is chosen at time t, and does not affect the firm's labour decision in following periods. The LP approach treats materials as a non-dynamic input, in the same way as it does labour, in a way that is consistent with a model where the materials and labour decisions are made simultaneously at time t.

The ACF approach treats the materials decision as made at time t, that is after the productivity shock realization, while the labour decision may be made at t, $t-1$, or $t-b$ where $0 < b < 1$. It is this timing difference that allows the ACF approach to turn the input demand function into a conditional demand function.

Assumption 4: scalar unobservable

The OP approach sets the firm's optimal investment decision as the solution to Equation (8), so that the firm's investment decision is consistent with the capital accumulation equation (7) and dependent only on the two state variables $k_{i,t}$ and $\omega_{i,t}$. This approach rules out capital adjustment costs between firms as well as differences in labour conditions or demand, while allowing for time variation (Ackerberg et al. 2015).

$$i_{i,t} = f_t(k_{i,t}, \omega_{i,t}) \tag{8}$$

The LP approach instead uses an unconditional demand function for intermediate inputs as in Equation (9), which is consistent with the case where the labour decision is made independently of the materials decision. ACF notes that this assumption requires for the underlying firms to operate in identical, or the same, labour and material input markets, and the same or identical output markets.[4]

$$m_{i,t} = f_t(k_{i,t}, \omega_{i,t}) \tag{9}$$

The ACF approach allows materials to be dependent on potentially dynamic labour as in Equation (10). That is the materials demand function in ACF is conditional on the labour input. The ACF, OP, and LP approach all rule out production functions with multiple structural unobservables.

$$m_{i,t} = \tilde{f}_t(k_{i,t}, l_{i,t}, \omega_{i,t}) \tag{10}$$

[4] ACF further notes that this assumption will also be satisfied if firms operate in the same output market with either homogeneous goods or completely symmetric product differentiation.

Assumption 5: strict monotonicity

The OP, LP, and ACF approaches require that their respective demand functions—Equations (8), (9), and (10)—are strictly increasing in $\omega_{i,t}$. This assumption is required to allow for inversion of the scalar unobservable, which allows for identification.

2.2 Estimation

This paper is concerned with the correction of simultaneity bias in inputs. We do not attempt to control for survival as in OP, as the unavailability of a firm in later periods may be due to late submissions. Survival may be controlled for by using the dormancy data in the panel available for firms in the post-2010 data. This field would likely only identify firms submitting tax returns despite dormancy status and exclude firms that failed to submit due to inactivity. The adequacy of this field has not been evaluated further and is beyond the scope of this paper. We rewrite the production as in ACF and LP where the production function is given by Equation (11).

$$q_{i,t} = \beta_0 + \beta_k k_{i,t} + \beta_l l_{i,t} + \omega_{i,t} + \varepsilon_{i,t} \tag{11}$$

Inverting the input demand function yields $\omega_{i,t} = \tilde{f}_t^{-1}(k_{i,t}, l_{i,t}, m_{i,t})$ for the ACF approach, so that we can rewrite the production function as in Equation (12). ACF and LP treat f_t^{-1} as a non-parametric function of $k_{i,t}$, $l_{i,t}$, and $m_{i,t}$, so that no coefficient can be separately identified in the resulting first stage moment condition given by Equation (13).

$$q_{i,t} = \beta_0 + \beta_k k_{i,t} + \beta_l l_{i,t} + \tilde{f}_t^{-1}(k_{i,t}, l_{i,t}, m_{i,t}) + \varepsilon_{i,t} = \tilde{\Phi}_t(k_{i,t}, l_{i,t}, m_{i,t}) + \varepsilon_{i,t+1} \tag{12}$$

$$E[\varepsilon_{i,t} | I_{i,t}] = E[q_{i,t} - \tilde{\Phi}_t(k_{i,t}, l_{i,t}, m_{i,t}) | I_{i,t}] \tag{13}$$

The first moment, Equation (13), does not allow for identification of $\tilde{\Phi}$ but provides an estimate $\hat{\tilde{\Phi}}$. The second moment, Equation (14), allows for identification of all parameters in the model, where $\tilde{\Phi}_{t-1}$ is replaced by its estimate from the first stage, where $\xi_{i,t}$ follows from the shock process definitions in Equations (3) and (4).

$$\begin{aligned} E[\xi_{i,t} + \varepsilon_{i,t} | I_{i,t-1}] = E\Bigg[& q_{i,t} - \beta_0 - \beta_k k_{i,t} - \beta_l l_{i,t} \\ & - g\Bigg(\tilde{\Phi}_{t-1}(k_{i,t-1}, l_{i,t-1}, m_{i,t-1}) \\ & - \beta_0 - \beta_k k_{i,t-1} - \beta_l l_{i,t-1}\Bigg) | I_{i,t-1}\Bigg] \end{aligned} \tag{14}$$

Following Ackerberg et al. (2015) we turn the conditional moment, Equation (14), into an unconditional moment as in Equation (15). We assume that the shock process is AR(1), so that $\xi_{i,t}$ is orthogonal to $\omega_{i,t-1}$, while allowing labour to be correlated to $\xi_{i,t}$. The four second stage moments to estimate β_0, β_k, β_l, and ρ are shown in Equation (15), where $l_{i,t}$ is instrumented by its lag to control for the correlation with $\xi_{i,t}$.

$$\begin{aligned} E\Bigg[\Bigg(& q_{i,t} - \beta_0 - \beta_k k_{i,t} - \beta_l l_{i,t} \\ & - \rho(\tilde{\Phi}_{t-1}(k_{i,t-1}, l_{i,t-1}, m_{i,t-1}) - \beta_0 - \beta_k k_{i,t-1} - \beta_l l_{i,t-1})\Bigg) \\ & \otimes \begin{pmatrix} 1 \\ k_{i,t} \\ l_{i,t-1} \\ \Phi_{t-1}(k_{i,t-1}, l_{i,t-1}, m_{i,t-1}) \end{pmatrix} \Bigg] = 0 \end{aligned} \tag{15}$$

2.3 Practical estimation

The first stage of the ACF approach regresses an interaction term of order N on $q_{i,t}$ and uses the prediction of this regression to yield $\hat{\hat{\Phi}}_t$. The shape of this polynomial in the ACF and the Rovigatti and Mollisi (2016) (PRODEST) implementations are defined as in Equation (16).

$$poly(x_1,...,x_N;P) = \begin{cases} 0 & \text{if } P=1 \\ \Sigma_{n_1=1}^{N} x_{n_1} \Sigma_{n_2=n_1}^{N} \beta_{n_1,n_2} x_{n_2} & \text{if } P=2 \\ \Sigma_{n_1=1}^{N} x_{n_1} \Sigma_{n_2=n_1}^{N} x_{n_2} \Sigma_{n_3=n_2}^{N} \beta_{n_1,n_2,n_3} x_{n_3} & \text{if } P=3 \\ ... & \\ \Sigma_{n_1=1}^{N} x_{n_1} \Sigma_{n_2=n_1}^{N} x_{n_2} \cdots \Sigma_{n_Q=n_{Q-1}}^{N} \beta_{n_1,n_2,...,n_Q} x_{n_Q} & \text{if } P=Q \end{cases} \tag{16}$$

The polynomial function is used as the measure for $\tilde{f}_t^{-1}$, in Equation (12), in the first stage regression function (17) estimated on the full sample.[5]

$$q_{i,t} = \beta_0 + \beta_l l_{i,t} + \beta_k k_{i,t} + \beta_m m_{i,t} + poly(l_{i,t}, k_{i,t}, m_{i,t}, P) \tag{17}$$

Two adjustments are made to the parameter space in the second stage, in order to reduce computational complexity in both the ACF and PRODEST approaches. The first simplification is to concentrate out the constant β_0 by using the $\hat{\hat{\Phi}}_{i,t}$ from the first stage to construct $\hat{\Omega}_{i,t}$ as in Equation (18), where β_l and β_k are the parameters to be estimated.

$$\hat{\Omega}_{i,t} = \widehat{\beta_0 + \omega_{i,t}} = \hat{\hat{\Phi}}_t(k_{i,t}, l_{i,t}, m_{i,t}) - \beta_k k_{i,t} - \beta_l l_{i,t} \tag{18}$$

The second simplification is then to regress $\hat{\Omega}_{i,t}$ on its own lag as in Equation (19). $\tilde{\Omega}_{i,t-1}$ is a matrix of the polynomial in $\hat{\Omega}_{t-1}$ and the constant. ACF uses a linear case where $\tilde{\Omega}_{i,t-1} = (1, \hat{\Omega}_{i,t-1})$, whereas the default option in PRODEST is a third order polynomial $\tilde{\Omega}_{i,t-1} = (1, \hat{\Omega}_{i,t-1}, \hat{\Omega}^2_{i,t-1}, \hat{\Omega}^3_{i,t-1})$. In Equation (19) ρ is the matrix for coefficients on $\tilde{\Omega}_{i,t-1}$.

$$\hat{\Omega}_{i,t} = \tilde{\Omega}_{i,t-1} \begin{pmatrix} \rho_0 \\ \rho_1 \\ \cdots \\ \rho_N \end{pmatrix} + \varepsilon_{i,t} = \tilde{\Omega}_{i,t-1}\rho + \varepsilon_{i,t} \tag{19}$$

The regression in Equation (19) is estimated using ordinary least squares (OLS) inside the generalized method of moments (GMM) estimator and is not minimized simultaneously.[6] The OLS assumptions then imply that the residuals from the regression are such that $E[\hat{\varepsilon}_{i,t}] = 0$, meaning that first and last moment conditions in Equation (15) hold. This reduces the parameter space of estimator to minimize the two moment conditions in Equation (20) where $Z_{i,t} = (k_{i,t}, l_{i,t-1})$.

$$E[Z'_{i,t}\hat{\xi}_{i,t}] = 0 \tag{20}$$

The GMM approach used in both PRODEST and the ACF procedure then finds β_k and β_l to minimize (21) where W is the weight matrix $W = (Z'Z/N)^{-1}$ as in the two stage least squares case (Wooldridge 2010).

$$J(\beta) = (Z'\xi)'W(Z'\xi) \tag{21}$$

[5] The full sample here refers to firms that satisfy the outlier constraints described in Section 4. The second stage regressions are limited to firms with valid lagged data.

[6] The estimation algorithms used in this paper, described in Section 2.4, do not make this adjustment.

2.4 Algorithms and implementation

We estimate the production function using the OLS, the Wooldridge (2009) GMM used in KN, the PRODEST implementation of the ACF routine, and our own implementation of the ACF routine.

A common concern with the ACF type estimators is convergence to spurious, but valid, global minima where $\hat{\beta}_l = \beta_k + \beta_l$ and $\hat{\beta}_k \to 0$.[7] Kim et al. (2019) find evidence that higher factor price dispersion may contribute to these biased estimates; specifically, they show that higher wage dispersion than the data-generating process considered by ACF may result in upward biased labour coefficients and downward biased capital coefficients. The PRODEST implementation appears to get around this issue by minimizing $J(\beta)$ using initial values from the OLS estimator and the Nelder-Mead decent algorithm with a step size of .00001. The combination of these initials and the small step size results in coefficients likely to converge to the OLS values if a local minima is present, but the algorithm can still converge to points outside the unit interval.

Our approach, which we call ACF(ρ), does not estimate the ρ matrix via OLS inside the GMM estimator, as in Equation (20), but expands the parameter space by minimizing $J(\beta,\rho)$ instead of $J(\beta)$. Our approach uses the starting values from the OLS approach for the β matrix as in the PRODEST and ACF's initial GAUSS implementation. We only specify non-zero initial values for ρ_0 and ρ_1 from Equation (22), where $\hat{\Omega}^{OLS}$ is the productivity estimates from the OLS initials. These initial values help ensure sensible AR(1) coefficients without introducing substantial computational complexity. Our approach also takes bigger initial steps than the PRODEST procedure and introduces noise only when taking larger steps.

$$\hat{\Omega}^{OLS}_{i,t} = \rho_0 + \rho_1 \hat{\Omega}^{OLS}_{i,t-1} \tag{22}$$

The code included with this paper uses Vega Yon and Quistorff (2019)'s parallelization routine in estimation to save computing time. Firms are split into their respective industries, after which the several algorithms are used on the varying samples.

3 Productivity and aggregation

We construct the log productivity measures for the OLS and Wooldridge (2009) GMM approaches using Equation (23). In the regressions using the PRODEST and ACF(ρ) algorithms, we estimate productivity using Equation (24).

$$\hat{\omega}_{i,t} = q_{i,t} - \hat{\beta}_l l_{i,t} - \hat{\beta}_k k_{i,t} \tag{23}$$

$$\hat{\omega}_{i,t} = \hat{\hat{\Phi}}_{i,t} - \hat{\beta}_l l_{i,t} - \hat{\beta}_k k_{i,t} \tag{24}$$

We construct the productivity aggregates using OP's approach using the definition in Equation (25). Productivity is aggregated by industry and year, weighing each firm by their respective sales share, as in Equation (26). In Equation (26) we normalize the industry-year aggregate of productivity to that of the first observed year. p_t, the industry's productivity aggregate, can be expanded to the unweighted mean of productivity and the sample covariance between productivity and output. The higher the covariance the more productivity is concentrated in more productive firms indicating higher allocative efficiency. In the results section we refer to $d_{i,t} = \sum_{i=1}^{N_t} \triangle s_{i,t} \triangle p_{i,t}$. We do not construct further aggregate indices as in Petrin and Levinsohn (2012).

[7] See ACF, footnote 16 (Ackerberg et al. 2015: 2438).

$$p_{i,t} = Exp(\hat{\omega}_{i,t}) \tag{25}$$

$$p_t = \frac{1}{p_0} \sum_{i=1}^{N_t} s_{i,t} p_{i,t} \tag{26}$$

$$p_t = \frac{1}{p_0} \left(\bar{p}_t + \sum_{i=1}^{N_t} \triangle s_{i,t} \triangle p_{i,t} \right) \tag{27}$$

4 Data

We use version 4.0 of the CIT-IRP5 panel (National Treasury and UNU-WIDER 2021). The main differences between this version of the data and the version used by KN is discussed in Ebrahim et al. (2021). We restrict the data to the 2009–17 sample due to well documented issues in the pre-2009 records and completeness issues in 2018. The panel is a combination of corporate income tax and individual pay-as-you-earn tax records. The primary source of data for this analysis is the corporate income tax data, and firms are required to self-declare income, expenditure, capital stock, and balance sheet information.

Version 4.0 of the data hosts significant improvements in industry classification. Budlender and Ebrahim (2020) harmonized existing industry variables and created a comparable set of classifications based on the Standard Industry Classification (SIC). Despite it being the most externally consistent industry measure, we do not use Budlender and Ebrahim (2020)'s composite profit code as it requires firms to have survived until 2013 to be assigned. We reclassify these industries to revision 4 of the International Standard Industry Classification as in KN. As our identification strategy requires estimation by industry, we assign firms to a single industry, even where they report to be operating in different industries over time. We assign a firm to manufacturing if it is observed in a manufacturing industry for 50 per cent or more of the years where it reports industry information. We assign these firms their mode industry; where multiple modes exists for a firm, we assign the firm to the latest manufacturing industry within which it was observed.[8]

Sales, cost of sales, and capital stock data come from the IT14 and ITR14 firm-level tax returns, while the employment data are aggregated by firm from IRP5 and IT3(a) individual tax returns, using the approach discussed in Ebrahim et al. (2021). The sales measure is the unadjusted sales measure from the respective forms, while our cost of sales measure corrects for opening and closing stocks. Value added is the difference between sales and the stock adjusted cost of sales measure. Value added is deflated by the value added deflator for manufacturing available from SARB (2021).[9] The construction of the capital stock measure represents a significant departure from KN and is discussed in detail below.

4.1 Capital stock

The property, plant, and equipment (PPE) measure is reported as a single variable for micro and small firms.[10] Medium-and-large firms report plant and equipment separately from property. PPE for these firms is calculated as the sum of the plant and equipment entry and the property entry. Medium-and-large firms also report other fixed assets as a separate field. Our fixed asset measure is the sum of the PPE

[8] Details on this assignment are available upon request.

[9] We use KBP6634, the value added deflator for manufacturing. The value added deflator is available at the quarterly level; we deflate each firm's value by the annualized deflator based on the quarter of the firm's end.

[10] See Ebrahim et al. (2021) for a detailed discussion on differences between the IT14 and ITR14 data in different vintages.

measure and other fixed assets.[11] When discussing PPE we refer to the property, plant, and equipment measure constructed in Ebrahim et al. (2021), which is the sum of property and plant and equipment in the case of medium-to-large firms, and when discussing fixed assets we refer to the sum of PPE and other fixed assets.

We construct a perpetual inventory capital stock measure, following the approach of Gal (2013) for European data. We assume that the capital accumulation is as in Equation (28), where $K_{i,t}$ refers to capital stock for firm i at time t. δ is the firm's depreciation rate in time t, and $I_{i,t}$ reflects the firm's investment during period t.

$$K_{i,t} = K_{i,t-1}(1-\delta_{i,t}) + I_{i,t} \tag{28}$$

In the construction of our measures, we ignore capital stock measures that are non-positive or missing.[12] We assume that the initial capital stock, $K_{i,0}$, is the first observed value for the firm. Investment is defined as the difference between capital stock and surviving capital stock from the previous period and is deflated by the capital formation deflator for manufacturing.[13] There is no direct measure of investment available in the CIT-IRP5 data.

$$I_{i,t} = \frac{K_{i,t} - (1-\delta_{i,t})K_{i,t-1}}{Deflator_{i,t}} \tag{29}$$

The IT14 and ITR14 forms allow all firms to declare a depreciation item in their expenses. This means firms of all sizes should have depreciation information. We construct the depreciation rate using present period depreciation expenditure as a ratio of previous period capital stock as in Equation (30). In Appendix A.1 we discuss the construction of four different depreciation rates based on the data, with and without imputations, as well as capital stock data with assumed depreciation rates of 10, 15, 20 per cent.

$$\delta_{i,t} = \frac{depreciation_{i,t}}{K_{i,t-1}} \tag{30}$$

In Appendix A.2 we show that the perpetual inventory fixed capital stock measure with imputations and assuming a depreciation rate of 10 per cent appears to be the most consistent measure of fixed capital stock, as it preserves most firms and matches the external data, based on the annual financial statistics (AFS), more consistently than other measures (StatsSA 2010–2018a).

4.2 Sample statistics

We require valid information on capital stock, number of employees, cost of sales, turnover, and the firm's industry in order to include it in our sample. While firms are required to submit capital stock, cost of sales, and turnover information, it appears that several firms report zero amounts. We treat fields with zero amounts or negative values as invalid and missing.

In Table 1 we show that only around 27 per cent of firms classified as manufacturers by Budlender and Ebrahim (2020) have sales, cost of sales, positive value added, capital stock, and employment data with the necessary lags and that satisfy sample restrictions. This proportion is consistent with that of KN. We lose a majority of observations, around 200,000, due to missing or zero sales figures. We lose around 30,000 firms due to invalid cost of sales data, and around 4,000 firms due to invalid value added

[11] k_ppe+k_faother in the data, as Ebrahim et al. (2021) incorporate the new vehicles field into PPE.

[12] These fields can be replaced with an imputed value. These imputed values are straight line imputations with a maximum length of two sequential years, as discussed in Appendices A and B.

[13] The deflator is the KBP6082 series available at the annual level in SARB (2021).

information. About 60,000 firms have invalid capital stock data and around 90,000 firms do not have valid employment data. We lose around another 50,000 firms due to invalid lagged data and another 25,000 due to sample restrictions. Our sample drops firms with an output to capital, output to labour, or capital-to-labour ratio above the 99th percentile or below the 1st percentile in any year from the sample. We drop these firms in all periods due to the lag structure of the estimator.

Table 1: Firms by data availability

	2010	2011	2012	2013	2014	2015	2016	2017	Total
Firms	71,181	71,884	75,436	76,132	77,093	77,600	78,355	75,259	662,628
with sales	50,318	50,172	51,194	51,554	51,837	51,776	52,256	51,114	457,907
	(70.69%)	(69.8%)	(67.86%)	(67.72%)	(67.24%)	(66.72%)	(66.69%)	(67.92%)	(69.1%)
and cost of sales	46,601	46,584	47,515	47,716	48,024	47,855	48,289	47,207	424,009
	(65.47%)	(64.8%)	(62.99%)	(62.68%)	(62.29%)	(61.67%)	(61.63%)	(62.73%)	(63.99%)
	[92.61%]	[92.85%]	[92.81%]	[92.56%]	[92.64%]	[92.43%]	[92.41%]	[92.36%]	[92.6%]
and positive	45,764	45,833	46,924	47,681	47,969	47,795	48,228	47,144	420,622
value added (VA)	(64.29%)	(63.76%)	(62.2%)	(62.63%)	(62.22%)	(61.59%)	(61.55%)	(62.64%)	(63.48%)
	[98.2%]	[98.39%]	[98.76%]	[99.93%]	[99.89%]	[99.87%]	[99.87%]	[99.87%]	[99.2%]
and capital stock	40,854	40,738	41,096	41,562	41,546	41,116	40,879	39,472	366,112
	(57.39%)	(56.67%)	(54.48%)	(54.59%)	(53.89%)	(52.98%)	(52.17%)	(52.45%)	(55.25%)
	[89.27%]	[88.88%]	[87.58%]	[87.17%]	[86.61%]	[86.03%]	[84.76%]	[83.73%]	[87.04%]
and employment	28,129	28,629	28,968	29,297	29,280	29,505	29,346	28,626	258,406
	(39.52%)	(39.83%)	(38.4%)	(38.48%)	(37.98%)	(38.02%)	(37.45%)	(38.04%)	(39%)
	[68.85%]	[70.28%]	[70.49%]	[70.49%]	[70.48%]	[71.76%]	[71.79%]	[72.52%]	[70.58%]
with lags	23,652	25,247	25,951	26,411	26,708	26,813	26,775	26,097	207,654
	(33.23%)	(35.12%)	(34.4%)	(34.69%)	(34.64%)	(34.55%)	(34.17%)	(34.68%)	(31.34%)
	[84.08%]	[88.19%]	[89.59%]	[90.15%]	[91.22%]	[90.88%]	[91.24%]	[91.17%]	[80.36%]
and sample	21,069	22,223	22,593	22,904	23,187	23,303	23,210	22,722	181,211
restrictions	(29.6%)	(30.92%)	(29.95%)	(30.08%)	(30.08%)	(30.03%)	(29.62%)	(30.19%)	(27.35%)
	[89.08%]	[88.02%]	[87.06%]	[86.72%]	[86.82%]	[86.91%]	[86.69%]	[87.07%]	[87.27%]

Note: this table shows the number of firms per year by data availability inclusive of the previous restriction. The number in round parentheses is the number of firms satisfying the restriction as a percentage of all firms in the manufacturing sector. The number in square brackets is the number of firms satisfying the restriction as a percentage of the number of firms satisfying the previous restriction.
Source: authors' calculations based on CIT-IRP5 version 4.0 (National Treasury and UNU-WIDER 2021) as discussed in (Ebrahim et al. 2021).

Table 2: Value added aggregates by data availability

	2010	2011	2012	2013	2014	2015	2016	2017
with sales	486,024	539,546	593,338	524,789	552,019	551,581	543,817	547,368
	90.36%	85.79%	90.66%	80.31%	81.72%	88.47%	88.36%	90.72%
and cost of sales	486,024	539,546	593,338	524,789	552,019	551,581	543,817	547,368
	[100%]	[100%]	[100%]	[100%]	[100%]	[100%]	[100%]	[100%]
	90.36%	85.79%	90.66%	80.31%	81.72%	88.47%	88.36%	90.72%
and capital stock	474,262	528,685	580,502	511,967	497,117	539,112	532,427	532,970
	(97.58%)	(97.99%)	(97.84%)	(97.56%)	(90.05%)	(97.74%)	(97.91%)	(97.37%)
	[97.58%]	[97.99%]	[97.84%]	[97.56%]	[90.05%]	[97.74%]	[97.91%]	[97.37%]
	88.17%	84.06%	88.7%	78.35%	73.59%	86.47%	86.51%	88.33%
and employment	409,490	453,911	495,120	459,395	451,658	466,482	466,118	446,574
	(84.25%)	(84.13%)	(83.45%)	(87.54%)	(81.82%)	(84.57%)	(85.71%)	(81.59%)
	[86.34%]	[85.86%]	[85.29%]	[89.73%]	[90.86%]	[86.53%]	[87.55%]	[83.79%]
	76.13%	72.17%	75.65%	70.3%	66.86%	74.82%	75.74%	74.01%
with lags	271,871	429,457	477,928	441,248	422,430	454,545	454,796	438,360
	(55.94%)	(79.6%)	(80.55%)	(84.08%)	(76.52%)	(82.41%)	(83.63%)	(80.09%)
	[66.39%]	[94.61%]	[96.53%]	[96.05%]	[93.53%]	[97.44%]	[97.57%]	[98.16%]
	50.54%	68.28%	73.03%	67.52%	62.54%	72.91%	73.9%	72.65%
and sample restrictions	231,828	366,757	411,192	373,355	362,179	388,131	387,604	377,293
	(47.7%)	(67.98%)	(69.3%)	(71.14%)	(65.61%)	(70.37%)	(71.27%)	(68.93%)
	[85.27%]	[85.4%]	[86.04%]	[84.61%]	[85.74%]	[85.39%]	[85.23%]	[86.07%]
	43.1%	58.32%	62.83%	57.13%	53.62%	62.26%	62.98%	62.53%
QFS value added	537,893	628,924	654,465	653,463	675,492	623,444	615,429	603,382

Note: this table shows total value added per year in the manufacturing sector by data availability in millions of 2012 ZAR. The percentage without parentheses is the aggregate of value added for firms satisfying the restriction as a proportion of the aggregate in the QFS data in the last row. The number in round parenthesis is the aggregate of value added for firms satisfying the restriction as a proportion of all firms with value added and valid sales data. The number in square brackets is the aggregate of value added for firms satisfying the restriction as a proportion of firms satisfying the previous restriction.
The value added measure in the QFS is turnover minus purchases plus closing stocks minus opening stock aggregated over the four quarters ending on Quarter 1 of the financial year in question. The figure for 2010 is 4× the QFS value for 2010. There is no data prior to Q1 of 2010 available in the QFS.
Source: authors' calculations based on CIT-IRP5 version 4.0 (National Treasury and UNU-WIDER 2021) as discussed in (Ebrahim et al. 2021).

In Table 2 we show that despite the apparent low representation of firms with the necessary sample, our coverage for value added is around 58 per cent of that reported in the quarterly financial statistics (QFS) (StatsSA 2010–2018b). KN reports a value added representation of around 25 per cent. In Table 3 we show that the capital stock aggregate by data availability similarly has around 50 per cent coverage compared to the AFS perpetual inventory measure, higher than the 23 per cent reported by KN. Our restrictions remove the substantial outliers in capital stock reported in 2010. In Table 4 our labour aggregate is shown to capture between 82 and 92 per cent of the quarterly employment statistics (QES) sample (StatsSA 2010–2018c). This increase in coverage compared to the value added and capital stock data is consistent with the increase in coverage reported in KN, which was also around 20–30 per cent higher than the coverage for the other values.[14]

[14] The relatively low representation in 2010 when requiring lags, compared to later years, is likely due to changes in the data vintages, as discussed in Ebrahim et al. (2021).

Table 3: Capital stock aggregates by data availability

	2010	2011	2012	2013	2014	2015	2016	2017
with sales	1,035,852	465,774	489,245	510,619	470,731	525,348	519,436	520,883
	291.3%	126.1%	107.1%	100.2%	93.14%	99.1%	99.08%	96.99%
and cost of sales	1,013,416	444,563	462,617	426,190	451,739	503,821	501,875	504,658
	[97.83%]	[95.45%]	[94.56%]	[83.47%]	[95.97%]	[95.9%]	[96.62%]	[96.89%]
	285%	120.4%	101.3%	83.6%	89.38%	95.04%	95.73%	93.97%
and positive VA	1,010,243	434,928	457,566	426,184	451,733	503,802	501,850	504,633
	(97.53%)	(93.38%)	(93.52%)	(83.46%)	(95.96%)	(95.9%)	(96.61%)	(96.88%)
	[99.69%]	[97.83%]	[98.91%]	[100%]	[100%]	[100%]	[100%]	[100%]
	284.1%	117.8%	100.2%	83.6%	89.38%	95.04%	95.72%	93.96%
and employment	912,589	330,614	347,874	355,670	379,181	384,360	392,955	395,254
	(88.1%)	(70.98%)	(71.1%)	(69.65%)	(80.55%)	(73.16%)	(75.65%)	(75.88%)
	[90.33%]	[76.02%]	[76.03%]	[83.45%]	[83.94%]	[76.29%]	[78.3%]	[78.33%]
	256.6%	89.53%	76.15%	69.76%	75.03%	72.5%	74.95%	73.59%
with lags	782,890	311,654	333,202	338,140	357,844	374,767	381,714	383,844
	(75.58%)	(66.91%)	(68.11%)	(66.22%)	(76.02%)	(71.34%)	(73.49%)	(73.69%)
	[85.79%]	[94.27%]	[95.78%]	[95.07%]	[94.37%]	[97.5%]	[97.14%]	[97.11%]
	220.1%	84.4%	72.94%	66.33%	70.81%	70.69%	72.81%	71.47%
and sample	120,418	220,258	239,604	238,657	254,429	274,335	284,526	293,716
restrictions	(11.63%)	(47.29%)	(48.97%)	(46.74%)	(54.05%)	(52.22%)	(54.78%)	(56.39%)
	[15.38%]	[70.67%]	[71.91%]	[70.58%]	[71.1%]	[73.2%]	[74.54%]	[76.52%]
	33.86%	59.65%	52.45%	46.81%	50.34%	51.75%	54.27%	54.69%
AFS capital stock	355,631	369,271	456,805	509,815	505,391	530,122	524,277	537,068

Note: this table shows total capital stock per year in the manufacturing sector by data availability in millions of 2012 ZAR using the perpetual inventory method. The percentage without parentheses is the aggregate of capital stock for firms satisfying the restriction as a proportion of the aggregate in the AFS data in the last row. The number in round parenthesis is the aggregate of capital stock for firms satisfying the restriction as a proportion of all firms with capital stock and valid sales data. The number in square brackets is the aggregate of capital stock for firms satisfying the restriction as a proportion of firms satisfying the previous restriction.
We discuss the construction of the perpetual inventory measure in Appendix A.
Source: authors' calculations based on CIT-IRP5 version 4.0 (National Treasury and UNU-WIDER 2021) as discussed in (Ebrahim et al. 2021).

Table 4: Labour aggregates by data availability

	2010	2011	2012	2013	2014	2015	2016	2017
with sales	1,268,773	1,320,085	1,354,885	1,362,092	1,382,172	1,388,377	1,390,191	1,409,539
	105.8%	112.6%	116%	116.9%	118.3%	119.6%	117.9%	118.3%
and cost of sales	1,181,332	1,244,331	1,275,593	1,265,330	1,299,702	1,309,589	1,315,555	1,338,806
	[93.11%]	[94.26%]	[94.15%]	[92.9%]	[94.03%]	[94.33%]	[94.63%]	[94.98%]
	98.53%	106.1%	109.2%	108.6%	111.2%	112.8%	111.6%	112.4%
and positive VA	1,165,227	1,220,118	1,264,563	1,265,307	1,299,413	1,309,531	1,315,199	1,338,550
	(91.84%)	(92.43%)	(93.33%)	(92.89%)	(94.01%)	(94.32%)	(94.61%)	(94.96%)
	[98.64%]	[98.05%]	[99.14%]	[100%]	[99.98%]	[100%]	[99.97%]	[99.98%]
	97.18%	104.1%	108.3%	108.6%	111.2%	112.8%	111.5%	112.4%
and capital stock	1,139,352	1,200,439	1,237,831	1,236,988	1,269,787	1,278,752	1,282,284	1,305,302
	(89.8%)	(90.94%)	(91.36%)	(90.82%)	(91.87%)	(92.1%)	(92.24%)	(92.6%)
	[97.78%]	[98.39%]	[97.89%]	[97.76%]	[97.72%]	[97.65%]	[97.5%]	[97.52%]
	95.03%	102.4%	106%	106.2%	108.7%	110.1%	108.7%	109.6%
with lags	889,980	1,128,923	1,180,642	1,182,523	1,219,486	1,237,987	1,246,471	1,268,174
	(70.14%)	(85.52%)	(87.14%)	(86.82%)	(88.23%)	(89.17%)	(89.66%)	(89.97%)
	[78.11%]	[94.04%]	[95.38%]	[95.6%]	[96.04%]	[96.81%]	[97.21%]	[97.16%]
	74.23%	96.28%	101.1%	101.5%	104.4%	106.6%	105.7%	106.4%
and sample restrictions	761,573	969,670	1,018,194	1,020,486	1,049,439	1,069,167	1,087,120	1,107,321
	(60.02%)	(73.46%)	(75.15%)	(74.92%)	(75.93%)	(77.01%)	(78.2%)	(78.56%)
	[85.57%]	[85.89%]	[86.24%]	[86.3%]	[86.06%]	[86.36%]	[87.22%]	[87.32%]
	63.52%	82.69%	87.19%	87.59%	89.81%	92.09%	92.19%	92.95%
QES employment	1,199,000	1,172,592	1,167,738	1,165,025	1,168,465	1,161,042	1,179,176	1,191,364

Note: this table shows total employment per year in the manufacturing sector by data availability in in weighted individuals. The percentage without parentheses is the aggregate of employment for firms satisfying the restriction as a proportion of the aggregate in the QES data in the last row. The number in round parenthesis is the aggregate of employment for firms satisfying the restriction as a proportion of all firms with employment and valid sales data. The number in square brackets is the aggregate of employment for firms satisfying the restriction as a proportion of firms satisfying the previous restriction.
The QES measure is the simple average of QES employment in manufacturing for the year in question; note that the QES measure is not weighted for part-time workers (StatsSA 2010–2018c).
Source: authors' calculations based on CIT-IRP5 version 4.0 (National Treasury and UNU-WIDER 2021) as discussed in (Ebrahim et al. 2021).

5 Results

In this section we discuss the results of the various estimators and implementations. The main coefficient estimates are reported in Appendix C. We provide the productivity estimates, coefficient estimates, and robustness results as Stata files.[15] We discuss the distribution of productivity by characteristics as in Kreuser and Newman (2018) and conclude by discussing productivity growth and the relationship between capital elasticities and allocative efficiency.

5.1 Coefficient estimates

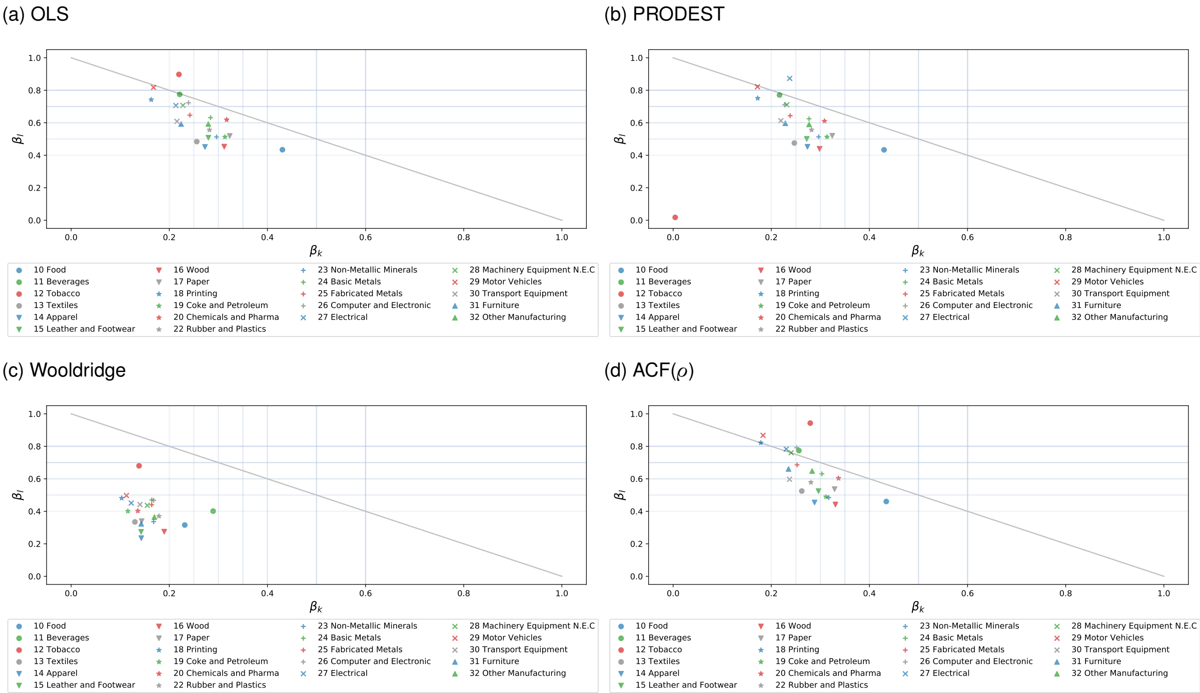
In Figure 1 we show the coefficients for the standard production function regression using the OLS, PRODEST, ACF, and GMM estimators. These estimates use the perpetual inventory capital stock measures with both imputed depreciation and a fixed depreciation rate of 10 per cent.[16] We generally find evidence of decreasing returns to scale for most industries with average capital and labour elasticities of around .27 and .63, respectively. Our estimates are in general higher than those found in the OLS and PRODEST approaches. The coefficients on the Wooldridge estimator are substantially lower than those of the direct ACF estimators; the Wooldridge estimator used by KN is based on the Petrin and Levin-

[15] Stata files available on the technical note's webpage as supplementary material.

[16] The GMM estimator is the same estimator used by Kreuser and Newman (2018).

sohn (2012) implementation and yields coefficients similar to the Wooldrdige implementation available in PRODEST. At present, the driving factor behind this decline in coefficients is not clear, as we do not appear to suffer from weak instruments.

Figure 1: Coefficients by estimator and type of depreciation for perpetual inventory fixed assets



Note: this figure shows the coefficients capital, β_k, and labour, β_l, at the the 2 digit ISIC4 industry classification for OLS in Panel 1a, PRODEST in Panel 1d, the Wooldridge implementation as in KN in Panel 1c, and our preferred algorithm conditioning optimizing over both the β and ρ matrices in Panel 1d. To simplify presentation we set $\beta > 1$ to $\beta = 1.25$ and $\beta < 0$ to $\beta = -.25$. Points to the right of the grey diagonal represent industries with increasing returns to scale, whereas lines to the left of the diagonal represent decreasing returns to scale.
The sample removes any firm that is observed to ever have a ratio between output and labour, output and capital, and capital and labour above the 99th percentile or below the 1st percentile within the industry.
Source: authors' calculations based on CIT-IRP5 version 4.0 (National Treasury and UNU-WIDER 2021) as discussed in (Ebrahim et al. 2021).

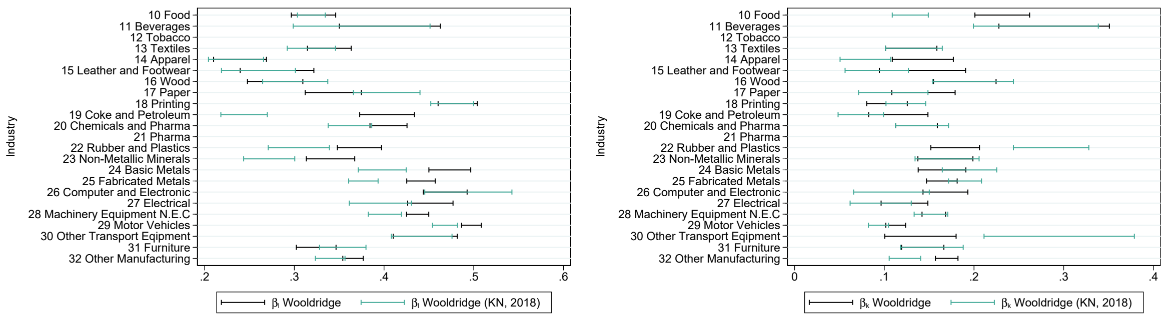
Mechanically, our ACF(ρ) implementation, discussed in Section 2.4 above, appears to provide reliable estimates more consistently than the PRODEST implementation for the CIT-IRP5 data. Our implementation does not converge to coefficients outside the unit interval for any industry except tobacco over all specifications of capital stock. The PRODEST converges to coefficients outside the unit interval for five industries when using perpetual inventory PPE with imputations, seen in Figure A5, and all but three industries are outside the unit interval for estimates using unadjusted PPE or fixed assets as the capital stock variable, as seen in Figure A7 in Appendix D. Our preferred approach does not converge to coefficients outside the unit interval using the perpetual inventory PPE measure, seen in Figure A5, and only converges to coefficients outside the unit interval for tobacco using the unadjusted measure. Furthermore, our approach does not fail to produce standard errors in any circumstance, whereas the PRODEST implementation fails to produce standard errors for two of the largest sectors: machinery and equipment N.E.C, and other manufacturing.

We compare our Wooldridge coefficients against those found by KN in Figure 2 by constructing a 5–95 per cent confidence interval using their standard errors. In Panel 2a our estimates for labour's elasticity are greater than those of KN for petroleum, rubber and plastics, basic metals, fabricated metals, electrical, machinery and equipment N.E.C, and motor vehicles. Our capital coefficient is above the 95 per cent confidence interval of KN's coefficient for the food, apparel, and other manufacturing sectors and below the interval for rubber and plastics, and other transport equipment. The capital measure used in

KN was the average real fixed capital stock for the firm in the current and previous year, meaning that their measure roughly approximated the perpetual inventory method. The fact that around 66 per cent of industries have overlapping confidence intervals for labour and that around 77 percent industries have overlapping capital coefficients shows that our estimates conform to KN's estimates despite dramatic changes in the industry classifications, labour variables, and capital stock variables.

Figure 2: Comparison to Wooldrige estimates in Kreuser and Newman (2018)

(a) β_l (b) β_k



Note: this figure shows the 95% confidence interval of the output elasticities of labour and capital, calculated as $\beta_i \pm 1.96 \times se_i$ for $i \in l,k$, for the Wooldridge estimator as reported in Appendix C, in black, compared to the same for the unweighted Wooldridge estimates reported in Kreuser and Newman (2018).

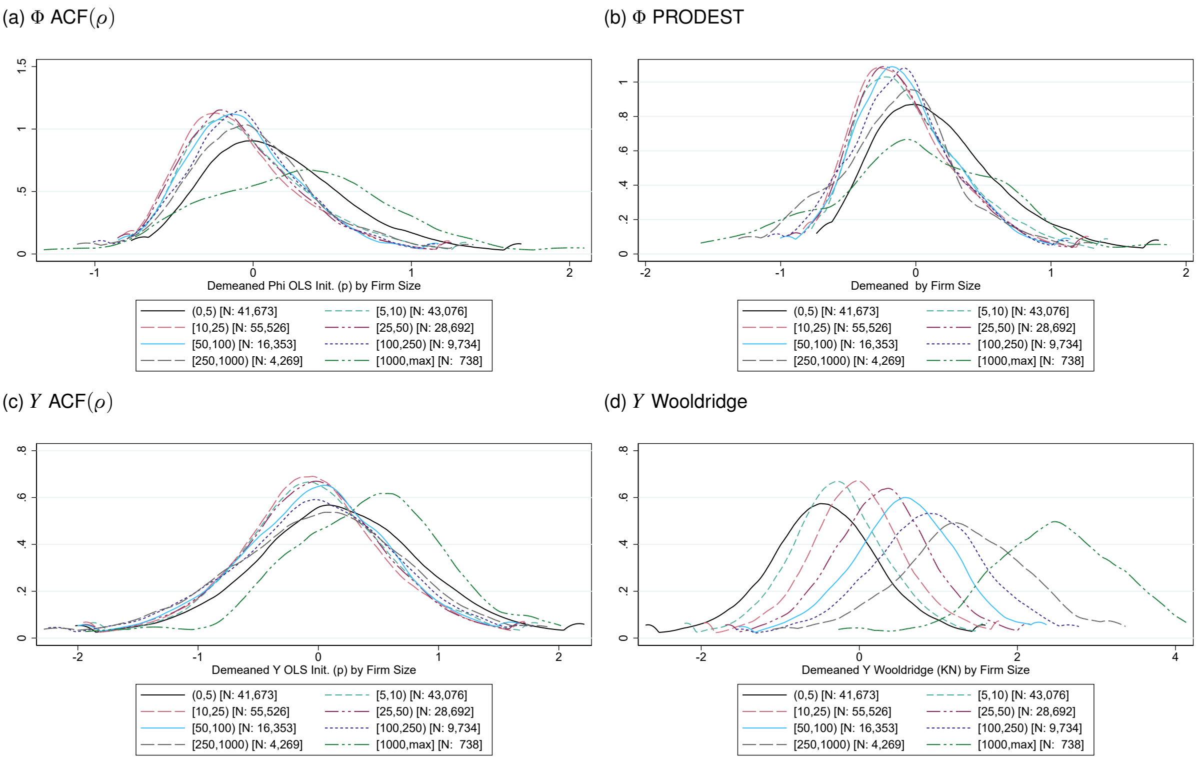
Source: authors' calculations based on CIT-IRP5 version 4.0 (National Treasury and UNU-WIDER 2021) as discussed in (Ebrahim et al. 2021).

5.2 Distribution

In Figure 3 we show the distribution of productivity by firm size for our preferred estimator, for the PRODEST estimator, for our preferred estimator using output instead of Φ, and for the Wooldridge estimator. All of the ACF type estimators show similar trends, with larger firms being generally more productive. The exception to this general trend is a relatively productive fringe of firms with less than five full-time employees, which show a mode around the general mean of all firm sizes and a substantial right tail. The distribution of firms with employees in the range of [5,10) appears to be significantly less productive than this group but still has a mode slightly higher than firms with employees in the range of [10,25). The productivity distribution for presumably more established firms, that is firms with 10 or more employees, follows the size dispersion found in KN more closely, with larger firms being more productive, and firms with a 1,000 or more employees being significantly more productive in general. In Panel 3d we find the smooth size distribution found in KN; as we do not weigh our productivity estimates by output, our distributions are tighter around the mean than theirs.

The substantial difference in the Wooldridge estimator's distribution and that of the ACF estimators can largely be attributed to the dramatically lower returns to scale implied by the former, as large portions of value added would be unaccounted for by the inputs and thus be assigned to employment in the distributions. The Wooldridge estimator further uses value added instead of Φ, the projected output from the control function, meaning that it potentially has more noise. In Panel 3c we show that the size premium in productivity is more pronounced when using output and not the projection in our preferred estimator. Our updated productivity estimates are broadly consistent with KN's estimates, but we do identify a competitive set of micro firms using our approach.

Figure 3: Distribution of TFP estimates by estimator and firm size



Note: this figure shows the distribution of total factor productivity, as discussed in Section 3. Each productivity measure is demeaned by both industry and year.

Source: authors' calculations based on CIT-IRP5 version 4.0 (National Treasury and UNU-WIDER 2021) as discussed in (Ebrahim et al. 2021).

5.3 Productivity growth and correlates

We provide the productivity indices of our preferred estimator in Table 5 for each sector. We find evidence of generally moderate productivity growth with only food, textiles, leather and footwear, wood, paper, and motor vehicles showing more than 5 per cent productivity growth since 2010. We do find constant or near declining productivity in apparel, printing, chemicals and pharmaceuticals, rubber and plastics, non-metallic minerals, fabricated metals, machinery and equipment N.E.C, other transport, and other manufacturing. The relatively high productivity growth for the chemical products and petroleum found by KN for 2010–13 is largely confirmed here, although the former industry appears to have had a negative productivity trajectory since 2015. Note that due to the updated industry indicators we cannot separate pharmaceuticals from general chemical products, but KN found evidence of declining productivity growth for pharmaceuticals. The wood and cork, fabricated metals, electrical equipment, and furniture sectors have low productivity concentrated in high output firms, with electrical equipment showing evidence of having high output assigned to unproductive firms. Some of these sectors have been investigated by the Competition Commission[17].

[17] See Bell et al. (2017) for a discussion on the fabricated metals and electrical equipment sectors and Competition Commission South Africa (2019c) for a discussion on the wood and cork manufacturers. Note that the Competition Commission has fined several cable and electric wire companies (Competition Commission South Africa 2019b). Furthermore, the fabricated metals sector likely includes AlcorMittal and related firms. In 2016 ArcelorMittal received what was at the time the highest penalty imposed on a single firm in the history of the Competition Commission. The commission and the firm, along with others, settled complaints including charges of information exchange, collusion, and excessive pricing that took place in the fabricated metals sector since at least 2003 (Competition Commission South Africa 2019a).

Table 5: TFP index for ACF(ρ)

	Industry	Agg.	2010	2011	2012	2013	2014	2015	2016	2017
10	Food	p_t	100.00	107.08	114.01	123.70	116.57	116.15	119.74	116.29
		$\bar{p}_t$	65.68	63.31	64.10	65.30	63.48	63.43	63.44	62.80
		d_t	34.32	43.77	49.91	58.40	53.10	52.72	56.31	53.49
11	Beverages	p_t	100.00	99.74	102.95	107.82	99.66	96.33	98.64	101.03
		$\bar{p}_t$	96.22	73.17	75.50	71.85	70.78	68.93	79.38	78.36
		d_t	3.78	26.57	27.45	35.97	28.88	27.40	19.26	22.67
12	Tobacco	p_t	100.00	92.85	87.35	98.12	113.17	102.60	102.72	126.17
		$\bar{p}_t$	98.69	93.17	96.60	92.48	86.99	81.80	83.22	94.48
		d_t	1.31	-0.32	-9.25	5.65	26.17	20.79	19.50	31.68
13	Textiles	p_t	100.00	104.87	110.62	105.04	103.40	101.17	108.52	101.84
		$\bar{p}_t$	86.79	87.98	86.38	87.03	86.26	85.99	86.21	82.46
		d_t	13.21	16.88	24.25	18.01	17.14	15.18	22.32	19.38
14	Apparel	p_t	100.00	97.93	110.17	104.77	107.25	102.45	101.77	98.98
		$\bar{p}_t$	73.83	75.14	74.01	73.42	74.46	72.51	72.48	70.60
		d_t	26.17	22.80	36.16	31.35	32.78	29.94	29.29	28.38
15	Leather and footwear	p_t	100.00	105.94	111.54	115.80	124.82	125.48	108.17	97.16
		$\bar{p}_t$	76.83	77.63	75.84	73.73	76.15	79.01	71.02	70.18
		d_t	23.17	28.32	35.69	42.07	48.68	46.46	37.15	26.97
16	Wood	p_t	100.00	98.82	100.61	102.82	101.26	102.48	105.86	99.53
		$\bar{p}_t$	86.63	88.76	88.30	89.41	88.77	89.12	87.25	87.46
		d_t	13.37	10.05	12.31	13.41	12.49	13.36	18.60	12.07
17	Paper	p_t	100.00	96.18	97.70	107.76	107.40	109.12	107.54	105.34
		$\bar{p}_t$	57.84	57.73	55.52	57.43	56.74	57.93	56.91	56.24
		d_t	42.16	38.45	42.19	50.33	50.67	51.19	50.63	49.10
18	Printing	p_t	100.00	104.66	108.29	107.62	95.54	102.71	104.06	96.95
		$\bar{p}_t$	79.89	80.24	79.65	78.99	77.73	78.13	79.19	78.37
		d_t	20.11	24.42	28.64	28.63	17.82	24.58	24.87	18.59
19	Coke and petroleum	p_t	100.00	102.09	104.68	103.68	108.02	109.80	98.37	101.12
		$\bar{p}_t$	79.20	78.32	80.84	80.66	79.71	78.02	73.99	74.30
		d_t	20.80	23.77	23.85	23.03	28.32	31.77	24.38	26.82
20	Chemicals and pharma	p_t	100.00	104.98	106.51	103.76	102.94	97.20	99.83	96.14
		$\bar{p}_t$	55.90	56.22	56.68	56.00	55.74	55.45	54.97	54.66
		d_t	44.10	48.76	49.83	47.76	47.20	41.76	44.86	41.47
22	Rubber and plastics	p_t	100.00	101.55	104.31	105.96	103.76	100.74	98.29	98.10
		$\bar{p}_t$	73.30	72.47	72.33	73.52	71.95	71.95	71.60	71.71
		d_t	26.70	29.08	31.98	32.43	31.81	28.78	26.69	26.38
23	Non-metallic minerals	p_t	100.00	91.73	96.23	96.32	99.89	100.34	97.36	97.40
		$\bar{p}_t$	50.44	50.72	51.37	51.34	50.68	50.72	49.50	48.90
		d_t	49.56	41.01	44.85	44.98	49.21	49.62	47.86	48.51
24	Basic metals	p_t	100.00	102.17	100.71	95.82	96.24	99.59	103.87	100.15
		$\bar{p}_t$	58.85	61.15	61.89	61.24	60.81	60.26	58.51	57.77
		d_t	41.15	41.03	38.82	34.59	35.43	39.33	45.37	42.37
25	Fabricated metals	p_t	100.00	102.59	102.94	100.28	98.94	100.46	98.29	97.01
		$\bar{p}_t$	82.88	84.50	85.22	84.50	83.50	83.08	81.82	81.86
		d_t	17.12	18.09	17.73	15.78	15.44	17.38	16.47	15.16

26	Computer and electronic	p_t	100.00	107.20	112.97	107.56	118.06	118.11	99.97	105.31
		$\bar{p}_t$	78.31	79.08	79.07	78.03	78.32	77.91	77.07	77.08
		d_t	21.69	28.13	33.90	29.54	39.74	40.20	22.90	28.23
27	Electrical	p_t	100.00	104.65	105.52	110.77	120.78	108.63	101.68	101.01
		$\bar{p}_t$	105.32	106.96	106.79	105.86	106.03	104.72	105.30	102.08
		d_t	-5.32	-2.31	-1.27	4.91	14.74	3.91	-3.62	-1.07
28	Machinery equipment N.E.C	p_t	100.00	107.41	116.35	113.82	106.21	100.98	100.05	98.54
		$\bar{p}_t$	79.17	79.86	80.65	81.01	79.70	78.07	76.97	76.70
		d_t	20.83	27.55	35.71	32.82	26.51	22.91	23.08	21.84
29	Motor vehicles	p_t	100.00	102.16	104.52	107.00	106.13	110.91	115.47	108.99
		$\bar{p}_t$	75.92	76.94	75.50	75.85	74.69	74.42	73.03	73.01
		d_t	24.08	25.21	29.02	31.16	31.43	36.48	42.44	35.99
30	Transport equipment	p_t	100.00	104.73	121.93	129.65	122.20	129.79	128.73	106.90
		$\bar{p}_t$	86.64	87.16	86.82	86.88	88.13	88.89	87.96	86.11
		d_t	13.36	17.57	35.12	42.77	34.07	40.90	40.76	20.79
31	Furniture	p_t	100.00	102.86	104.94	107.92	110.76	109.29	104.37	104.34
		$\bar{p}_t$	91.71	91.86	90.82	92.85	93.71	92.50	90.18	90.98
		d_t	8.29	11.00	14.12	15.06	17.05	16.79	14.18	13.36
32	Other manufacturing	p_t	100.00	100.46	105.31	106.33	107.18	107.41	102.61	98.81
		$\bar{p}_t$	79.04	79.96	80.05	79.95	79.47	78.29	77.61	76.44
		d_t	20.96	20.50	25.26	26.38	27.72	29.12	24.99	22.37

Note: this table shows the productivity aggregates using the ACF (ρ) algorithm and aggregating according to Equation (27). p_t is the aggregate for the industry as a whole, $\bar{p}_t$ is the unweighted average productivity estimate in each industry, and $d_t = \sum_{i=1}^{N_t} \triangle s_{i,t} \triangle p_{i,t}$ is the sample covariance between productivity and output. The share used is sales. We remove outliers in each industry.

Source: authors' calculations based on CIT-IRP5 version 4.0 (National Treasury and UNU-WIDER 2021) as discussed in (Ebrahim et al. 2021).

In Table 5, d_t is the sample covariance between output and productivity. The higher this measure the more output is concentrated in highly productive firms in an industry. We find that changes in this covariance explains the majority of changes in productivity in the given sectors. Food, paper, chemicals and pharmaceuticals, and non-metallic minerals are the industries that appear to have more output concentrated in highly productive firms. These industries are also those with the lowest ratio of labour elasticity to capital elasticity, as seen in Appendix C. In Table 6 we regress d_t on the coefficients of the various estimators, as well as the aggregate capital-to-labour, output-to-capital, and output-to-labour indices.[18] We consistently find a large and significant relationship between the capital elasticity of output to the index, indicating that industries with higher allocative efficiency are more productive in their use of capital. The fact that the coefficient on labour is weakly positive in the ACF type estimators appears to indicate that these firms are better in their use of labour as well. We find evidence that industries with higher allocative efficiency are also those with higher aggregate capital-to-labour ratios and higher aggregate output-to-capital ratios. Taken together, we interpret these results as strengthening KN's results that higher capital–labour ratio firms are more productive in the sense that industries with better allocation are also those that use capital more efficiently.

[18] The indices are constructed by using the aggregate ratio of the sample in 2010 for each industry. We do not use 2010 data in these regressions as the capital-to-labour indices will always equal unity.

Table 6: Correlation between covariance of productivity and output to factor input intensity and elasticities

	ACF (ρ)	PRODEST	Wooldridge (KN)
$\hat{\beta}_k$	1.63***	1.13***	.487*
	(.257)	(.239)	(.266)
$\hat{\beta}_l$	.221**	.012	-.02
	(.111)	(.106)	(.150)
Aggregate capital-to-labour index	.064***	.069***	.091***
	(.018)	(.019)	(.022)
Aggregate output-to-labour index	-.04	-.07**	-.07*
	(.033)	(.033)	(.038)
Aggregate output-to-capital index	.006***	.006***	.005**
	(.001)	(.001)	(.002)
Constant	-.33**	-.01	.201**
	(.143)	(.122)	(.091)
Adj. R^2	.353	.288	.109
Observations	147	147	147

Note: this table shows the regression coefficients, where the dependent variable is $d_t = \sum_{i=1}^{N_t} \triangle s_{i,t} \triangle p_{i,t}$. $\hat{\beta}_k$ and $\hat{\beta}_l$ are the regression coefficients from the estimator named in the column. The aggregate indices are calculated as $AggregateIndex_{i,t} = \frac{X_{i,t}/Y_{i,t}}{X_{i,2010}/Y_{i,2010}}$. Tobacco and 2010 data are excluded from the sample.
Source: authors' calculations based on CIT-IRP5 version 4.0 (National Treasury and UNU-WIDER 2021) as discussed in (Ebrahim et al. 2021).

6 Conclusion

This paper updated the productivity estimates for the South African manufacturing sector first documented in Kreuser and Newman (2018) by extending the time frame and utilizing updated employment, capital stock, and industry variables. We further provide perpetual inventory capital stock data that approximates the trends seen in the annual financial statistics (StatsSA 2010–2018a).

We expand KN's approach by using the Wooldridge (2009) and Ackerberg et al. (2015) approaches. We present the ACF estimates from Rovigatti and Mollisi (2016) and our own implementation, showing that our implementation converges to coefficients inside the unit interval more consistently. Our approach appears to be more consistent to changes in the sample and starting values.

We find limited productivity growth in South African manufacturing from 2010 to 2017 with the majority of within-industry variation over time being due to changes in allocative efficiency. The study notes that the food, paper, and chemicals and pharmaceuticals sectors consist of high-output and highly productive firms. Industries that are highly allocative efficient tend to employ relatively more capital to labour and by the same token have a larger proportion of output attributed to capital. The fabricated metals, wood and cork, electrical equipment, and furniture sectors, on the other hand, display evidence of poor allocative efficiency. Finally, we find capital intensity and generally better use of inputs to be positively correlated with allocative efficiency.

References

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). 'Identification Properties of Recent Production Function Estimators'. *Econometrica*, 83(6): 2411–51. https://doi.org/10.3982/ECTA13408

Bell, J., Nhundu, N., Paelo, A., Thosago, M., and Vilakazi, T. (2017). 'Growth and Strategies of Large and Leading firms-Metals, Machinery and Equipment Sector Assessment'. CCRED Working Paper 18/2017. Johannesburg: Centre for Competition, Regulation and Economic Development, University of Johannesburg. https://doi.org/10.2139/ssrn.3013675

Budlender, J., and Ebrahim, A. (2020). 'Industry Classification in the South African Tax Microdata'. WIDER Working Paper 2020/99. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2020/856-6

Collard-Wexler, A., and De Loecker, J. (2020). 'Production Function Estimation and Capital Measurement Error'. NBER Working Paper 22437. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w22437

Competition Commission South Africa (2019a). *Annual Report 2016-17.* Pretoria: Competition Commission South Africa. Available at: https://www.compcom.co.za/wp-content/uploads/2019/10/Annual-Report-2016-17.pdf (accessed 28 September 2021)

Competition Commission South Africa (2019b). *Annual Report 2017-18: Competition Regulation for a Growing and Inclusive Economy.* Pretoria: Competition Commission South Africa. Available at: https://www.compcom.co.za/wp-content/uploads/2019/10/annual-report-2017-2018.pdf (accessed 3 September 2021)

Competition Commission South Africa (2019c). 'Wood-Based Panel Companies Prosecuted for Cartel Conduct'. Media Statement. Available at: https://www.compcom.co.za/wp-content/uploads/2020/01/Wood-based-panel-companies-prosecuted-for-cartel-conduct.pdf (accessed 3 September 2021)

Ebrahim, A., Kreuser, C. F., and Kilumelume, M. (2021). 'The Guide to the CIT-IRP5 Panel Version 4.0'. WIDER Working Paper 2021/173. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2021/113-6

Gal, P. N. (2013). 'Measuring Total Factor Productivity at the Firm Level Using OECD-ORBIS'. OECD Economics Department Working Paper 1049. Paris: OECD. https://doi.org/10.1787/5k46dsb25ls6-en

Kerr, A. (2020). 'Earnings in the South African Revenue Service IRP5 Data'. WIDER Working Paper 2020/62. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2020/819-1

Kim, K. i., Luo, Y., and Su, Y. (2019). 'A Robust Approach to Estimating Production Functions: Replication of the ACF Procedure'. *Journal of Applied Econometrics*, 34(4): 612–19. https://doi.org/10.1002/jae.2697

Kreuser, C. F., and Newman, C. (2018). 'Total Factor Productivity in South African Manufacturing Firms'. *South African Journal of Economics*, 86(S1): 40–78. https://doi.org/10.1111/saje.12179

Levinsohn, J., and Petrin, A. (2003). 'Estimating Production Functions Using Inputs to Control for Unobservables'. *The Review of Economic Studies*, 70(2): 317–41. https://doi.org/10.1111/1467-937X.00246

National Treasury and UNU-WIDER (2021). 'CIT-IRP5 Panel 2008–2018 [dataset]'. Version 4.0. Pretoria: South African Revenue Service [producer of the original data], 2020. Pretoria: National Treasury and UNU-WIDER [producer and distributor of the harmonized dataset], 2021.

Olley, G. S., and Pakes, A. (1996). 'The Dynamics of Productivity in the Telecommunications Equipment Industry'. *Econometrica*, 64(6): 1263–97. https://doi.org/10.2307/2171831

Petrin, A., and Levinsohn, J. (2012). 'Measuring Aggregate Productivity Growth Using Plant-Level Data'. *The RAND Journal of Economics*, 43(4): 705–25. https://doi.org/10.1111/1756-2171.12005

Rovigatti, G., and Mollisi, V. (2016). 'PRODEST: Stata Module for Production Function Estimation Based on the Control Function Approach'. Statistical Software Component S458239. Chestnut Hill: Department of Economics, Boston College. (revised 28 July 2020)

SARB (2021). 'National Accounts March 2021'. Quarterly Bulletin. Available at: https://www.resbank.co.za/content/dam/sarb/publications/quarterly-bulletins/download-information-from-xlsx-data-files/2021/march-2021/05Kbp6NationalAccountsMarch2021.zip (accessed 6 September 2021)

StatsSA (2010–2018a). *P0021 Annual Financial Statistics (AFS).* Pretoria: Statistics South Africa. Available at: http://www.statssa.gov.za?page_id=1866&PPN=P0021&SCH=7846 (accessed 6 September 2021)

StatsSA (2010–2018b). *P0044 Quarterly Financial Statistics (QFS): March, June, September, December.* Pretoria: Statistics South Africa. Available at: http://www.statssa.gov.za?page_id=1866&PPN=P0044&SCH=72763 (accessed 6 September 2021)

StatsSA (2010–2018c). *P0277 Quarterly Employment Statistics (QES): March, June, September, December.* Pretoria: Statistics South Africa. Available at: http://www.statssa.gov.za?page_id=1866&PPN=P0277&SCH=7900 (accessed 6 September 2021)

Vega Yon, G. G., and Quistorff, B. (2019). 'parallel: A Command for Parallel Computing'. *The Stata Journal*, 19(3): 667–84. https://doi.org/10.1177/1536867X19874242

Wooldridge, J. M. (2009). 'On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables'. *Economics letters*, 104(3): 112–14. https://doi.org/10.1016/j.econlet.2009.04.026

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (2nd ed.). Cambridge, MA: MIT Press.

A Capital stock

The most common approach to measure capital stock in the CIT-IRP5 panel is to sum property, plant, and equipment and other fixed assets. Firms often switch between filling in the property, plant, and equipment field and the other fixed assets field over time. This measure is still generally noisy, as seen in the coefficient estimates in Appendix D. This appendix discusses the construction of perpetual inventory capital stock measures along with other capital stock measures.

Tables A1 and A2 show the completeness of PPE and fixed assets for the manufacturing sector from 2009 to 2017. There are on average 50,000 manufacturing firms with fixed assets per year, whereas the number of firms reporting PPE is around 40,000 up to 2012, after which it also becomes around 50,000. This shift may be due to the shift from IT14 to ITR14 in this period. Independent of the capital stock measure, around 97 per cent of firms with capital data also have sales data, with these firms accounting for around 98–99 per cent of total capital stock in the period. Firms with positive value added have capital stock accounting for around 95 per cent of the aggregate and 90 per cent of observations.[19] While there is a significant drop in the aggregate capital stock for firms with positive value added in 2013—with the aggregate only accounting for 83 per cent of the unrestricted sample—once restricting the sample to firms with positive employment data, representation appears more consistent around an average of 70 per cent of the aggregate despite some deviations in 2009, 2010, and 2014. The number of manufacturing firms with positive value added and employment averages to around 31,000 firms per year, whereas the number of firms also reporting depreciation data averages to around 30,000 firms per year. Firms with positive depreciation data account for around 95 per cent of firms with employment data up to 2012 and around 93 per cent of firms from 2013 onwards.

The aggregates show substantial outliers in the raw data for 2009 and 2010[20]. For the manufacturing subsample these entries appear to be true outliers and do not appear to shift dramatically when conditioning on additional variables.

[19] Positive value added here means that the firms have positive sales and cost of sales amounts, with the cost of sales amount being lower than the sales amount.

[20] In later figures we censor these outliers.

Table A1: PPE for manufacturing

Aggregate	2009	2010	2011	2012	2013	2014	2015	2016	2017
PPE	158,035	842,002	311,411	350,262	367,635	359,850	421,295	448,705	465,451
	37,215	38,619	38,574	42,749	52,140	51,810	51,075	51,151	49,680
with sales	151,740	832,509	306,921	338,232	358,121	354,035	417,239	444,854	459,980
	(96.02%)	(98.87%)	(98.56%)	(96.57%)	(97.41%)	(98.38%)	(99.04%)	(99.14%)	(98.82%)
	36,049	37,385	37,354	41,556	50,965	50,734	50,031	50,090	48,692
	[96.87%]	[96.8%]	[96.84%]	[97.21%]	[97.75%]	[97.92%]	[97.96%]	[97.93%]	[98.01%]
and cost of sales	145,555	822,524	297,416	328,259	304,794	343,044	402,587	429,966	450,263
	(95.92%)	(98.8%)	(96.9%)	(97.05%)	(85.11%)	(96.9%)	(96.49%)	(96.65%)	(97.89%)
	33,667	34,835	34,933	38,793	47,452	47,402	46,726	46,888	45,666
	[93.39%]	[93.18%]	[93.52%]	[93.35%]	[93.11%]	[93.43%]	[93.39%]	[93.61%]	[93.79%]
and pos. VA	138,898	819,926	289,533	323,182	304,785	343,035	402,352	429,947	450,244
	(95.43%)	(99.68%)	(97.35%)	(98.45%)	(100%)	(100%)	(99.94%)	(100%)	(100%)
	33,063	34,314	34,484	38,387	47,436	47,382	46,702	46,866	45,650
	[98.21%]	[98.5%]	[98.71%]	[98.95%]	[99.97%]	[99.96%]	[99.95%]	[99.95%]	[99.96%]
% of total	87.89%	97.38%	92.97%	92.27%	82.9%	95.33%	95.5%	95.82%	96.73%
% of obs.	88.84%	88.85%	89.4%	89.8%	90.98%	91.45%	91.44%	91.62%	91.89%
and employment	119,935	774,914	220,719	249,531	253,941	290,218	302,432	339,718	355,018
	(86.35%)	(94.51%)	(76.23%)	(77.21%)	(83.32%)	(84.6%)	(75.17%)	(79.01%)	(78.85%)
	23,319	23,919	24,394	26,572	32,296	32,310	32,397	32,477	31,915
	[70.53%]	[69.71%]	[70.74%]	[69.22%]	[68.08%]	[68.19%]	[69.37%]	[69.3%]	[69.91%]
% of total	75.89%	92.03%	70.88%	71.24%	69.07%	80.65%	71.79%	75.71%	76.27%
% of obs.	62.66%	61.94%	63.24%	62.16%	61.94%	62.36%	63.43%	63.49%	64.24%
and depreciation	117,403	774,288	219,528	246,778	249,374	280,734	292,576	331,974	344,413
	(97.89%)	(99.92%)	(99.46%)	(98.9%)	(98.2%)	(96.73%)	(96.74%)	(97.72%)	(97.01%)
	22,571	23,138	23,483	25,353	30,341	30,288	30,207	30,123	29,424
	[96.79%]	[96.73%]	[96.27%]	[95.41%]	[93.95%]	[93.74%]	[93.24%]	[92.75%]	[92.19%]
% of total	74.29%	91.96%	70.49%	70.46%	67.83%	78.01%	69.45%	73.98%	74%
% of obs.	60.65%	59.91%	60.88%	59.31%	58.19%	58.46%	59.14%	58.89%	59.23%

Note: this table shows the aggregates and number of observations available for PPE for manufacturing from 2009–17 by data availability. The first row is the aggregate over all firms with positive data in the year, in millions, without additional restrictions. The second row is the number of observations with positive available data without additional restrictions.
The remainder of the rows provide the same information conditioning on the availability of positive sales, cost of sales, positive value added, employment, and depreciation data. The number in round parentheses is the aggregate for the firms satisfying the criteria as a percentage of the aggregate in the previous criteria. The number in square brackets is the number of firms satisfying the criteria as a percentage of the firms satisfying the previous criteria. The '% of total' is the aggregate for firms satisfying the criteria as a percentage of the aggregate for all firms with positive values without additional restrictions. The '% of obs.' is the number of firms satisfying the criteria as a percentage of the number of firms with positive values without additional restrictions.
Source: authors' calculations based on CIT-IRP5 version 4.0 (National Treasury and UNU-WIDER 2021) as discussed in (Ebrahim et al. 2021).

Table A2: Fixed for manufactring

Aggregate	2009	2010	2011	2012	2013	2014	2015	2016	2017
Fixed	675,944	912,549	372,598	422,037	442,037	428,549	505,920	525,835	551,327
	46,919	48,479	47,374	49,124	53,064	52,566	51,909	51,872	50,334
with sales	668,225	901,531	367,155	408,547	431,254	420,931	499,390	521,645	544,939
	(98.86%)	(98.79%)	(98.54%)	(96.8%)	(97.56%)	(98.22%)	(98.71%)	(99.2%)	(98.84%)
	45,469	46,942	45,890	47,736	51,859	51,457	50,833	50,791	49,322
	[96.91%]	[96.83%]	[96.87%]	[97.17%]	[97.73%]	[97.89%]	[97.93%]	[97.92%]	[97.99%]
and cost of sales	661,345	890,440	356,206	396,419	363,346	407,143	482,373	504,185	532,619
	(98.97%)	(98.77%)	(97.02%)	(97.03%)	(84.25%)	(96.72%)	(96.59%)	(96.65%)	(97.74%)
	42,352	43,662	42,827	44,523	48,311	48,098	47,495	47,563	46,260
	[93.14%]	[93.01%]	[93.33%]	[93.27%]	[93.16%]	[93.47%]	[93.43%]	[93.64%]	[93.79%]
and pos. VA	654,198	887,527	348,137	391,050	363,338	407,134	482,136	504,160	532,588
	(98.92%)	(99.67%)	(97.73%)	(98.65%)	(100%)	(100%)	(99.95%)	(100%)	(99.99%)
	41,638	43,007	42,270	44,051	48,294	48,077	47,469	47,540	46,242
	[98.31%]	[98.5%]	[98.7%]	[98.94%]	[99.96%]	[99.96%]	[99.95%]	[99.95%]	[99.96%]
% of total	96.78%	97.26%	93.43%	92.66%	82.2%	95%	95.3%	95.88%	96.6%
% of obs.	88.74%	88.71%	89.23%	89.67%	91.01%	91.46%	91.45%	91.65%	91.87%
and employment	139,258	808,014	258,387	293,867	299,361	338,496	358,785	390,867	411,456
	(21.29%)	(91.04%)	(74.22%)	(75.15%)	(82.39%)	(83.14%)	(74.42%)	(77.53%)	(77.26%)
	28,606	29,284	29,385	30,497	33,045	32,917	33,062	33,075	32,429
	[68.7%]	[68.09%]	[69.52%]	[69.23%]	[68.42%]	[68.47%]	[69.65%]	[69.57%]	[70.13%]
% of total	20.6%	88.54%	69.35%	69.63%	67.72%	78.99%	70.92%	74.33%	74.63%
% of obs.	60.97%	60.41%	62.03%	62.08%	62.27%	62.62%	63.69%	63.76%	64.43%
and depreciation	136,580	807,227	257,042	290,866	294,157	326,265	345,909	381,223	399,427
	(98.08%)	(99.9%)	(99.48%)	(98.98%)	(98.26%)	(96.39%)	(96.41%)	(97.53%)	(97.08%)
	27,620	28,227	28,193	29,059	31,057	30,865	30,845	30,674	29,892
	[96.55%]	[96.39%]	[95.94%]	[95.28%]	[93.98%]	[93.77%]	[93.29%]	[92.74%]	[92.18%]
% of total	20.21%	88.46%	68.99%	68.92%	66.55%	76.13%	68.37%	72.5%	72.45%
% of obs.	58.87%	58.23%	59.51%	59.15%	58.53%	58.72%	59.42%	59.13%	59.39%

Note: this table shows the aggregates and number of observations available for Fixed for manufacturing from 2009–17 by data availability. The first row is the aggregate over all firms with positive data in the year, in millions, without additional restrictions. The second row is the number of observations with positive available data without additional restrictions.
The remainder of the rows provide the same information conditioning on the availability of positive sales, cost of sales, positive value added, employment, and depreciation data. The number in round parentheses is the aggregate for the firms satisfying the criteria as a percentage of the aggregate in the previous criteria. The number in square brackets is the number of firms satisfying the criteria as a percentage of the firms satisfying the previous criteria. The '% of total' is the aggregate for firms satisfying the criteria as a percentage of the aggregate for all firms with positive values without additional restrictions. The '% of obs.' is the number of firms satisfying the criteria as a percentage of the number of firms with positive values without additional restrictions.
Source: authors' calculations based on CIT-IRP5 version 4.0 (National Treasury and UNU-WIDER 2021) as discussed in (Ebrahim et al. 2021).

A.1 Depreciation

The depreciation data itself is problematic, as there is good reason to believe that firms over-report their depreciation rates in the data since they can potentially write it off as an expense. The QFS and AFS datasets consistently report an aggregate depreciation rate of 10 per cent, whereas the aggregate depreciation rate in the manufacturing sector is around 13 per cent. This aggregate hides substantial firm-level variation, as shown in Figure A1, where the median depreciation rate for manufacturing is around 30 per cent after censoring outliers.

We construct eight depreciation rates using the depreciation data provided in the CIT-IRP5 data.

- We create four main depreciation rates based on the data:
  1. The first reflects the definition in Equation (30). That is, we define the depreciation rate to be given by current depreciation over the previous period's capital stock.

2. The second ensures that reported depreciation for the current period is less than capital stock in the previous period.

3. The third uses the first variable but implements a one-year forward imputation if missing.

4. The last uses the second but implements a one-year forward imputation if missing.

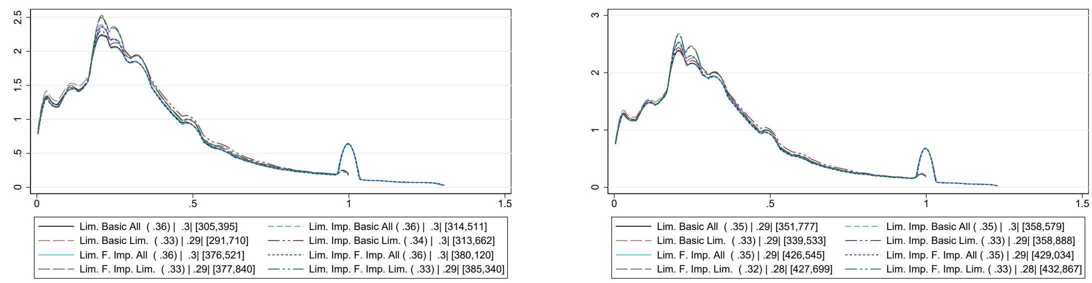
- The other four depreciation variables use the same definition as those above but use an imputed value where missing. The imputation method is discussed in Appendix B.

In Figure A1 we show the distribution of depreciation rates for PPE and fixed assets. The censoring of the data is crucial for sensible coefficients as even after controlling for the 1 per cent and 99 per cent, the data is censored to be below 1.5 for meaningful distributions. As seen, the imputation approaches of capital stock do not significantly alter the distribution of depreciation rates in any of the distribution except to reduce the proportion of firms reporting 1 or greater depreciation rates.

Figure A1: Depreciation rates in CIT-IRP5 data based on capital stock variable and imputations

(a) PPE

(b) Fixed assets



Note: this figure shows the distribution of depreciation rates for manufacturing firms based on Equation (30) for PPE in Panel A1a and for fixed assets in Panel A1b. In all cases we restrict our sample to firms with positive depreciation and positive lagged capital stock. We do not report the figures for firms with a depreciation rate of greater than 150 per cent (1.5) for clarity. Basic All uses the depreciation rate as provided by the data without additional limits, Basic Lim. limits the depreciation so that $\delta < 1$. F. Imp. All uses the next period's depreciation rate if the current period's depreciation rate is missing. F. Imp. Lim. uses the forward depreciation rate as long as $\delta < 1$.

Source: authors' calculations based on CIT-IRP5 version 4.0 (National Treasury and UNU-WIDER 2021) as discussed in (Ebrahim et al. 2021).

In addition to using reported depreciation, we assume a depreciation rate of 10, 15, and 20 per cent. The capital stock value used for these deprecation rates are: $K^D_{i,t} = K_{i,t} + Depreciation^{BV}_{i,t}$. We treat the book value of depreciation as 0 where missing or negative. We then specify the $\delta_{i,t}$ for use in constructing the perpetual inventory values.

A.2 Preferred capital stock measures

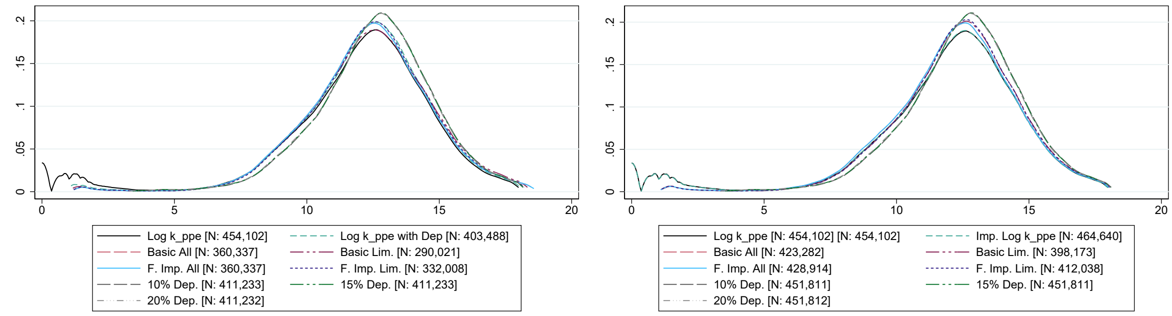
In Figure A2 we show the distribution on Log PPE and Log Imputed PPE for the manufacturing sector. Limiting PPE to firms with positive depreciation shows a general tightening of the distribution and a loss of around 50,000 observations mostly at the lower end of the distribution.[21]. The perpetual inventory measure without any imputations loses around 90,000 observations. Imputing PPE values but not further imputing missing or invalid depreciation values results in a loss of an additional 70,000 observations. Using forward imputation for depreciation values in the latter case saves 40,000 of these observations. The perpetual inventory methods with assumed depreciation lose around 40,000 observations. When using imputed capital stock measures, we gain around 10,000 observations by not restricting to firms with depreciation data. The forward imputation of depreciation and PPE results in 60,000 more ob-

[21] Note these figures are for the entire period 2008–18

servations than in the case of the unlimited depreciation forward imputation in the non-imputed capital stock, and 80,000 observations in the limited series. The full imputed limited series, F. Imp. Lim. in Panel (b), is arguably the most appropriate measure to use when wishing to base your capital stock measures on observed depreciation for as many observations as possible; this measure does not report capital stock for around 40,000 observations, whereas the non-imputed series loses around 120,000 observations. Using assumed depreciation without imputed capital stock results in a loss of around 40,000 observations, whereas assumed depreciation with imputed capital stock results in a loss of around 3,000 observations.

Figure A2: Distribution of perpetual inventory PPE based on depreciation imputation

(a) No PPE imputations (b) With PPE and depreciation imputations



Note: this figure shows the distribution of Log PPE for firms in manufacturing sector by depreciation data. Subfigure (a) shows the distribution of PPE without imputations, whereas (b) shows the distribution of PPE with imputations, as discussed in Appendix B. The number in square brackets is the number of observations with strictly positive data. Data above the 99 per cent and below the 1 per cent are removed for clarity of presentation.

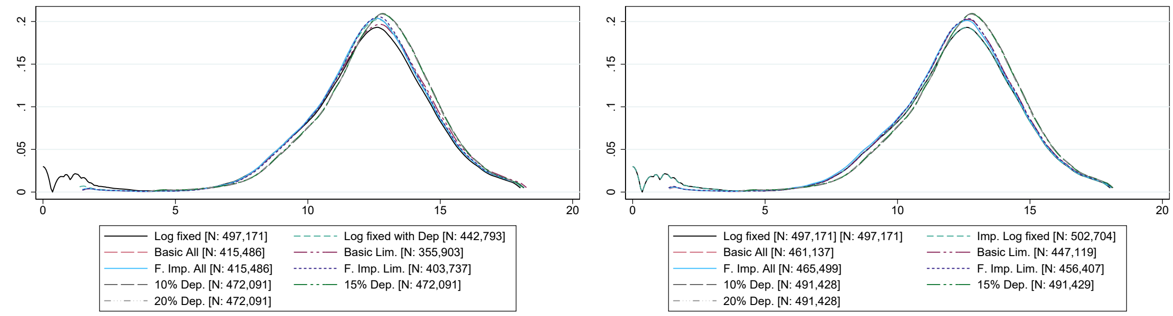
In (a) Log k_ppe is the non-perpetual inventory measure of PPE and Log k_ppe with Dep. is the same measure limited to firms with depreciation data. In (b) Log k_ppe is the same as in (a), but Imp. Log k_ppe is the number of distribution of the imputed capital variable. Ideally, Log k_ppe and Imp Log k_ppe should follow the same distribution.

Basic All refers to perpetual inventory capital stock data using the depreciation rate, as provided by the data without limits. Basic Lim. limits the depreciation so that $\delta < 1$. F. Imp. All uses the next period's depreciation rate if the current period's depreciation rate is missing. F. Imp. Lim. uses the forward depreciation rate, as long as $\delta < 1$. The X% Dep. sets assume that the depreciation rate is X%.

Source: authors' calculations based on CIT-IRP5 version 4.0 (National Treasury and UNU-WIDER 2021) as discussed in (Ebrahim et al. 2021).

Figure A3: Distribution of perpetual inventory fixed assets based on depreciation imputation

(a) No fixed imputations (b) With fixed and depreciation imputations



Note: this figure shows the distribution of Log fixed assets for firms in manufacturing sector by depreciation data. Subfigure (a) shows the distribution of fixed assets without imputations, whereas (b) shows the distribution of fixed assets with imputations, as discussed in Appendix B. The number in square brackets is the number of observations with strictly positive data. Data above the 99 per cent and below the 1 per cent are removed for clarity of presentation.

In (a) Log fixed is the non-perpetual inventory measure of fixed assets and Log fixed with Dep. is the same measure limited to firms with depreciation data. In (b) Log fixed is the same as in (a), but Imp. Log fixed is the number of distribution of the imputed capital variable. Ideally, Log fixed and Imp Log fixed should follow the same distribution.

Basic All refers to perpetual inventory capital stock data using the depreciation rate, as provided by the data without limits. Basic Lim. limits the depreciation so that $\delta < 1$. F. Imp. All uses the next period's depreciation rate if the current period's depreciation rate is missing. F. Imp. Lim. uses the forward depreciation rate, as long as $\delta < 1$. The X% Dep. sets assume that the depreciation rate is X%.

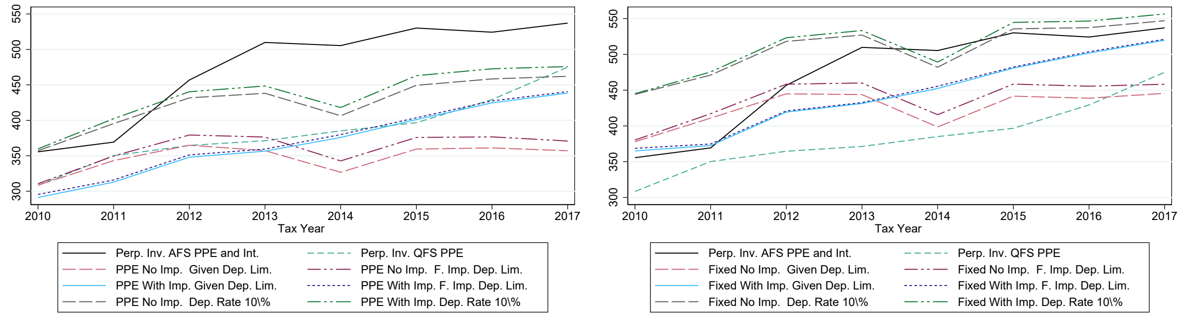
Source: authors' calculations based on CIT-IRP5 version 4.0 (National Treasury and UNU-WIDER 2021) as discussed in (Ebrahim et al. 2021).

The distribution of the perpetual inventory measures in manufacturing generally appears to tighten at the lower end. The basic and forward unlimited imputation generally has a lower peak than the limited series but follows a broadly similar distribution. The perpetual inventory methods using assumed depreciation rates appear to be a significantly tighter distribution with a higher mean. The forward imputed limited set is extremely similar to the distribution of log k_ppe limited to firms with depreciation data. The distributional impact of the varying depreciation measures are similar for the perpetual inventory PPE measures with capital and depreciation imputations, where the main difference in distributions are based on whether a perpetual inventory approach is used or not, or whether real depreciation information is used. In Figure A3 we show the distributions of log fixed assets and log fixed assets with and without imputations. The results are broadly similar to that of Figure A3.

In Figure A4 we show the time series aggregates of the perpetual inventory fixed assets measure for the manufacturing sector. As seen, the perpetual inventory method with an assumed depreciation rate of 10 per cent most consistently matches the AFS fixed asset data. The imputed approach is further preferred as it preserves most observations.

Figure A4: Perpetual inventory fixed assets compared to AFS and QFS

(a) Perpetual inventory PPE (b) Perpetual inventory fixed assets



Note: this figure shows the aggregate perpetual inventory capital stock measures for the AFS, QFS, and CIT-IRP5 data for the PPE and Fixed assets measures. The AFS value includes intangible assets. No Imp. indicates that the capital stock measure was not imputed where missing. With Imp. indicates that the capital stock was imputed where missing or invalid. Given Dep. indicates that the depreciation rate was used as given. F. Imp. indicates that the forward imputed value was used when depreciation was missing. Dep. Rate 10% indicates that the deprecation rate was assumed to be 10%.

Source: authors' calculations based on CIT-IRP5 version 4.0 (National Treasury and UNU-WIDER 2021) as discussed in (Ebrahim et al. 2021).

B Smoothing

We impute data for selected capital stock and the depreciation rate following the approach of Petrin and Levinsohn (2012) used to impute data for firms that exit and enter the sample.[22] The smoothing approach essentially attempts to fit a straight line through two points. While there are arguably several more sophisticated ways to solve the imputation problem,[23] the present approach is preferred for at least three reasons.

Firstly, the imputation will never yield an imputed value that is outside of the given distribution of the firm. That is, the approach will never lead to an imputed value that is higher or lower than the values reported by the firm. Secondly, the approach allows for flexible shocks, although it will necessarily reduce variation.[24] Finally, by using a local smoother based only on the variable in question, we do not force structure on the data based on economically important information in the other fields. While this may seem counter intuitive, it does allow for counter-acting shocks to follow in the firm's structure. We do limit the amount of data imputed, however, by only allowing data to be used within a two-period range in each direction and not imputing any data before the first date or after the last date of available data. Note that since up to two periods are used to impute interior data, the data are weighed by proximity to the date in question. This allows for two points to be imputed smoothly. The smoothing algorithm is only executed once.

$$imputed\ variable_t = \begin{cases} variable_t & if \quad variable \in \mathbb{R}_{>0} \\ imp(variable_t) & if \quad variable_t \notin \mathbb{R}_{>0}\ and\ imp(variable_t) \in \mathbb{R}_{>0} \end{cases} \tag{31}$$

$$\begin{aligned} imp(variable_t) =& variable_{[t-n|available]} \times \left(\frac{[t+n|available]-t}{[t+n|available]-[t-n|available]} \right) \\ &+ variable_{[t+n|available]} \times \left(\frac{t-[t-n|available]}{[t+n|available]-[t-n|available]} \right) \end{aligned}$$

[22] The Petrin and Levinsohn (2012) approach is not used in estimation, but only to calculate aggregate productivity movements.

[23] In fact, there is an entire multiple imputation literature, machine learning literature, and so on. Note that several fixed effect approaches have also been attempted.

[24] Regression-based approaches using full firm level data will smooth shocks over the entire period, the present approach will only smooth shocks over two periods at most. Note that the local smoother, as opposed to a smoother based on firm, industry, and year effects, will also ignore potentially small within-firm sample issues or the assignment of values that are not appropriate for the firm in question. A high-growth firm will have a relatively meaningless fixed effect, as it would be on a mean that may never have occurred (recall that for each firm the mean is technically a draw from a very small sample $N \in [0,10]$). Then, using a fixed effect by firm, industry, and year may result in substantial imputed shocks at either the beginning or end of the period, depending on where the missing data exists.

C Coefficients by industry and estimator

Table A3: Estimates, part 1

(a) 10 Food

Estimator	β_l	β_k	ρ_1	Stats
OLS*	.434***	.431***		.8
	(9.1e-03)	(6.4e-03)		7,273
Wooldridge (KN)*	.316***	.232***		.86
	(9.7e-03)	(.016)		7,273
PRODEST ACF**	.433***	.43***		.***
	(.08)	(.09)		7,273
ACF$(\rho)^\dagger$	.46***	.434***	.878***	.0014
	(4.5e-03)	(8.4e-04)	(6.8e-06)	7,273

(b) 11 Beverages

Estimator	β_l	β_k	ρ_1	Stats
OLS*	.775***	.221***		.76***
	(.021)	(.013)		2,312
Wooldridge (KN)*	.401***	.289***		.84
	(.026)	(.031)		2,312
PRODEST ACF**	.771***	.217***		.***
	(.03)	(.054)		2,312
ACF$(\rho)^\dagger$	.774***	.256***	.905***	.0029
	(5.0e-03)	(2.8e-03)	(3.8e-05)	2,312

(c) 12 Tobacco

Estimator	β_l	β_k	ρ_1	Stats
OLS*	.898***	.22***		.79
	(.046)	(.03)		345
Wooldridge (KN)*	.68***	.138**		.81
	(.068)	(.059)		345
PRODEST ACF**	.017	4.0e		.***
	(4.01)	(.975)		345
ACF$(\rho)^\dagger$	.943***	.279***	.843***	1.4e-09
	(8.7e-03)	(7.2e-03)	(7.1e-03)	345

(d) 13 Textiles

Estimator	β_l	β_k	ρ_1	Stats
OLS*	.484***	.256***		.7
	(9.8e-03)	(6.4e-03)		5,213
Wooldridge (KN)*	.335***	.13***		.8
	(.01)	(.015)		5,213
PRODEST ACF**	.475***	.247***		.
	(1.2e-03)	(6.3e-03)		5,213
ACF$(\rho)^\dagger$	.525***	.262***	.895***	5.3e-04
	(4.4e-03)	(6.1e-04)	(2.1e-06)	5,213

(e) 14 Apparel

Estimator	β_l	β_k	ρ_1	Stats
OLS*	.451***	.273***		.62
	(.012)	(9.5e-03)		3,704
Wooldridge (KN)*	.235***	.143***		.76
	(.013)	(.017)		3,704
PRODEST ACF**	.452***	.274***		.
	(.029)	(6.6e-03)		3,704
ACF$(\rho)^\dagger$	.454***	.288***	.91***	7.6e-04
	(5.2e-03)	(1.7e-03)	(2.4e-06)	3,704

(f) 15 Leather and footwear

Estimator	β_l	β_k	ρ_1	Stats
OLS*	.508***	.28***		.7
	(.017)	(.012)		1,978
Wooldridge (KN)*	.274***	.143***		.81
	(.017)	(.025)		1,978
PRODEST ACF**	.501***	.272***		.
	(.088)	(.055)		1,978
ACF$(\rho)^\dagger$	.525***	.296***	.904***	.0016
	(6.6e-03)	(2.5e-03)	(1.6e-05)	1,978

(g) 16 Wood

Estimator	β_l	β_k	ρ_1	Stats
OLS*	.453***	.312***		.7
	(.012)	(8.7e-03)		3,632
Wooldridge (KN)*	.275***	.19***		.8
	(.014)	(.018)		3,632
PRODEST ACF**	.44***	.298***		.
	(.017)	(.016)		3,632
ACF$(\rho)^\dagger$	.442***	.331***	.901***	8.0e-04
	(2.7e-03)	(1.0e-03)	(2.9e-06)	3,632

(h) 17 Paper

Estimator	β_l	β_k	ρ_1	Stats
OLS*	.519***	.323***		.77
	(.013)	(8.5e-03)		3,491
Wooldridge (KN)*	.339***	.144***		.85
	(.014)	(.018)		3,491
PRODEST ACF**	.52***	.324***		.
	(1.6e-03)	(8.4e-03)		3,491
ACF$(\rho)^\dagger$	.536***	.329***	.906***	6.4e-04
	(5.3e-03)	(1.3e-03)	(3.1e-06)	3,491

Note: this table shows the coefficient estimates for the varying estimators for each ISIC4 industry at the 2-digit level. The sample removes firms that ever had a ratio of output to labour, output to capital, or capital to labour lower than the 1st percentile or greater than the 99th percentile for the entire industry for each individual year. β_l and β_k are the coefficients for labour and capital, respectively. ρ_1, where available, is the coefficient on the prediction of lag productivity at the current parameters as in Equation (19). The Stats column shows the relevant fit statistic in the first row and the number of observations in the second row. The standard errors, in parentheses, for the PRODEST and other ACF estimators are obtained from bootstrapping with 1,000 replications. The p-values for these estimators are obtained from a two-tailed test on the normal distribution.
ACF(ρ) refers to the procedure minimizing $J(\beta,\rho)$. * The relevant fit statistics is the adjusted-R^2. ** The PRODEST ACF procedure does not produce fit statistics by default; where the standard errors on the PRODEST estimator are zero the bootstrapping did not produce them. † The relevant fit statistic for the ACF(ρ) algorithms is $J(.)$ from Equation (21). In the β and ρ columns: * $p < .1$, ** $p < .05$, and *** $p < .01$.
In the stats columns: * $p > .05$ and ** $p > .1$ for the null-hypothesis that $\beta_l + \beta_k = 1$.
Source: authors' calculations based on CIT-IRP5 version 4.0 (National Treasury and UNU-WIDER 2021) as discussed in (Ebrahim et al. 2021).

Table A4: Estimates, part 2

(a) 18 Printing

Estimator	β_l	β_k	ρ_1	Stats
OLS*	.742***	.163***		.76
	(8.4e-03)	(5.5e-03)		8,116
Wooldridge (KN)*	.481***	.103***		.83
	(.011)	(.012)		8,116
PRODEST ACF**	.751***	.172***		.*
	(9.5e-03)	(.037)		8,116
ACF(ρ)†	.822***	.179***	.828***	6.4e-04***
	(1.1e-03)	(1.9e-04)	(3.1e-06)	8,116

(b) 19 Coke and petroleum

Estimator	β_l	β_k	ρ_1	Stats
OLS*	.513***	.313***		.63
	(.013)	(7.7e-03)		3,314
Wooldridge (KN)*	.401***	.116***		.74
	(.014)	(.017)		3,314
PRODEST ACF**	.513	.314**		.***
	(.363)	(.1)		3,314
ACF(ρ)†	.49***	.312***	.875***	.0021
	(2.7e-04)	(1.4e-04)	(6.1e-06)	3,314

(c) 20 Chemicals and pharma

Estimator	β_l	β_k	ρ_1	Stats
OLS*	.619***	.317***		.75
	(9.2e-03)	(6.1e-03)		8,573
Wooldridge (KN)*	.402***	.136***		.85
	(9.2e-03)	(.012)		8,573
PRODEST ACF**	.61***	.309***		.
	(1.3e-05)	(5.9e-06)		8,573
ACF(ρ)†	.603***	.337***	.918***	8.7e-04
	(2.9e-03)	(9.8e-04)	(2.7e-06)	8,573

(d) 22 Rubber and plastics

Estimator	β_l	β_k	ρ_1	Stats
OLS*	.556***	.282***		.73
	(9.8e-03)	(6.3e-03)		6,762
Wooldridge (KN)*	.37***	.179***		.81
	(.011)	(.014)		6,762
PRODEST ACF**	.556***	.282***		.
	(4.1e-06)	(3.9e-07)		6,762
ACF(ρ)†	.578***	.281***	.894***	5.8e-04
	(4.1e-03)	(4.3e-04)	(1.7e-06)	6,762

(e) 23 Non-metallic minerals

Estimator	β_l	β_k	ρ_1	Stats
OLS*	.513***	.296***		.66
	(.011)	(7.2e-03)		5,901
Wooldridge (KN)*	.337***	.168***		.78
	(.012)	(.016)		5,901
PRODEST ACF**	.513***	.297***		.
	(1.7e-05)	(2.8e-06)		5,901
ACF(ρ)†	.485***	.317***	.918***	.0011
	(2.4e-06)	(1.1e-03)	(2.0e-06)	5,901

(f) 24 Basic metals

Estimator	β_l	β_k	ρ_1	Stats
OLS*	.632***	.284***		.73
	(9.7e-03)	(6.4e-03)		7,503
Wooldridge (KN)*	.47***	.164***		.82
	(.01)	(.014)		7,503
PRODEST ACF**	.624***	.277***		.
	(8.2e-07)	(2.7e-07)		7,503
ACF(ρ)†	.631***	.303***	.904***	5.8e-04
	(4.6e-03)	(5.3e-04)	(2.2e-06)	7,503

(g) 25 Fabricated metals

Estimator	β_l	β_k	ρ_1	Stats
OLS*	.647***	.242***		.71
	(5.9e-03)	(3.9e-03)		16,099
Wooldridge (KN)*	.44***	.164***		.77
	(7.5e-03)	(8.7e-03)		16,099
PRODEST ACF**	.643***	.239***		.
	(5.2e-07)	(1.7e-06)		16,099
ACF(ρ)†	.686***	.253***	.852***	3.0e-04
	(2.4e-03)	(2.6e-06)	(1.7e-06)	16,099

(h) 26 Computer and electronic

Estimator	β_l	β_k	ρ_1	Stats
OLS*	.723***	.239***		.74
	(.01)	(6.5e-03)		5,986
Wooldridge (KN)*	.468***	.168***		.81
	(.012)	(.013)		5,986
PRODEST ACF**	.711***	.227***		.
	(3.2e-07)	(3.1e-06)		5,986
ACF(ρ)†	.792***	.252***	.852***	7.1e-04
	(2.7e-03)	(4.3e-04)	(4.3e-06)	5,986

Note: this table shows the coefficient estimates for the varying estimators for each ISIC4 industry at the 2-digit level. The sample removes firms that ever had a ratio of output to labour, output to capital, or capital to labour lower than the 1st percentile or greater than the 99th percentile for the entire industry for each individual year. β_l and β_k are the coefficients for labour and capital, respectively. ρ_1, where available, is the coefficient on the prediction of lag productivity at the current parameters as in Equation (19). The Stats column shows the relevant fit statistic in the first row and the number of observations in the second row. The standard errors, in parentheses, for the PRODEST and other ACF estimators are obtained from bootstrapping with 1,000 replications. The p-values for these estimators are obtained from a two-tailed test on the normal distribution.
ACF(ρ) refers to the procedure minimizing $J(\beta, \rho)$. * The relevant fit statistics is the adjusted-R^2. ** The PRODEST ACF procedure does not produce fit statistics by default; where the standard errors on the PRODEST estimator are zero the bootstrapping did not produce them. † The relevant fit statistic for the ACF(ρ) algorithms is $J(.)$ from Equation (21). In the β and ρ columns: * $p < .1$, ** $p < .05$, and *** $p < .01$.
In the stats columns: * $p > .05$ and ** $p > .1$ for the null-hypothesis that $\beta_l + \beta_k = 1$.
Source: authors' calculations based on CIT-IRP5 version 4.0 (National Treasury and UNU-WIDER 2021) as discussed in (Ebrahim et al. 2021).

Table A5: Estimates, part 3

(a) 27 Electrical

Estimator	β_l	β_k	ρ_1	Stats
OLS*	.706***	.213***		.77
	(.01)	(7.2e-03)		5,535
Wooldridge (KN)*	.451***	.123***		.84
	(.013)	(.013)		5,535
PRODEST ACF**	.873***	.238***		.***
	(.088)	(.04)		5,535
ACF$(\rho)^\dagger$	.783***	.231***	.847***	2.7e-04
	(2.9e-03)	(3.1e-04)	(3.9e-06)	5,535

(b) 28 Machinery equipment N.E.C

Estimator	β_l	β_k	ρ_1	Stats
OLS*	.707***	.228***		.73
	(5.3e-03)	(3.6e-03)		22,658
Wooldridge (KN)*	.437***	.155***		.82
	(5.9e-03)	(6.7e-03)		22,658
PRODEST ACF**	.711	.232		.
	(0)	(0)		22,658
ACF$(\rho)^\dagger$	.76***	.24***	.856***	5.7e-04***
	(2.0e-03)	(2.1e-05)	(4.9e-06)	22,658

(c) 29 Motor vehicles

Estimator	β_l	β_k	ρ_1	Stats
OLS*	.817***	.168***		.79
	(4.6e-03)	(2.8e-03)		23,085
Wooldridge (KN)*	.497***	.113***		.87
	(5.5e-03)	(5.6e-03)		23,085
PRODEST ACF**	.821***	.172***		.
	(4.5e-07)	(2.3e-06)		23,085
ACF$(\rho)^\dagger$	.868***	.183***	.897***	.001
	(9.2e-04)	(6.7e-05)	(2.2e-06)	23,085

(d) 30 Other transport equipment

Estimator	β_l	β_k	ρ_1	Stats
OLS*	.609***	.216***		.66
	(.014)	(9.2e-03)		3,011
Wooldridge (KN)*	.442***	.14***		.75
	(.016)	(.02)		3,011
PRODEST ACF**	.613***	.22***		.
	(.034)	(.016)		3,011
ACF$(\rho)^\dagger$	.597***	.237***	.897***	6.2e-04
	(5.3e-03)	(1.2e-03)	(2.3e-06)	3,011

(e) 31 Furniture

Estimator	β_l	β_k	ρ_1	Stats
OLS*	.592***	.224***		.69
	(8.8e-03)	(6.5e-03)		6,222
Wooldridge (KN)*	.323***	.143***		.79
	(.011)	(.012)		6,222
PRODEST ACF**	.597***	.229***		.
	(.042)	(7.9e-03)		6,222
ACF$(\rho)^\dagger$	.66***	.235***	.837***	4.3e-04
	(3.2e-03)	(2.1e-04)	(4.1e-06)	6,222

(f) 32 Other manufacturing

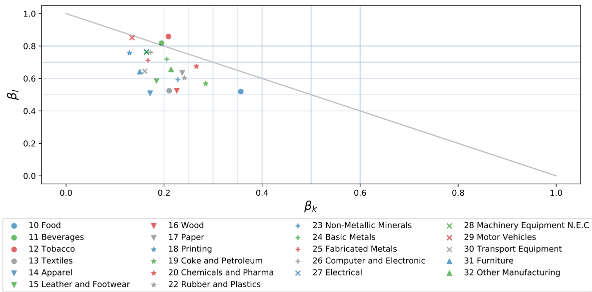
Estimator	β_l	β_k	ρ_1	Stats
OLS*	.592***	.28***		.73
	(4.5e-03)	(3.0e-03)		29,899
Wooldridge (KN)*	.364***	.17***		.81
	(5.2e-03)	(6.4e-03)		29,899
PRODEST ACF**	.59	.277		.
	(0)	(0)		29,899
ACF$(\rho)^\dagger$	.648***	.283***	.876***	4.7e-04
	(2.2e-03)	(1.8e-05)	(5.7e-06)	29,899

Note: this table shows the coefficient estimates for the varying estimators for each ISIC4 industry at the 2-digit level. The sample removes firms that ever had a ratio of output to labour, output to capital, or capital to labour lower than the 1st percentile or greater than the 99th percentile for the entire industry for each individual year. β_l and β_k are the coefficients for labour and capital, respectively. ρ_1, where available, is the coefficient on the prediction of lag productivity at the current parameters as in Equation (19). The Stats column shows the relevant fit statistic in the first row and the number of observations in the second row. The standard errors, in parentheses, for the PRODEST and other ACF estimators are obtained from bootstrapping with 1,000 replications. The p-values for these estimators are obtained from a two-tailed test on the normal distribution.
ACF(ρ) refers to the procedure minimizing $J(\beta,\rho)$. * The relevant fit statistics is the adjusted-R^2. ** The PRODEST ACF procedure does not produce fit statistics by default; where the standard errors on the PRODEST estimator are zero the bootstrapping did not produce them. † The relevant fit statistic for the ACF(ρ) algorithms is $J(.)$ from Equation (21). In the β and ρ columns: * $p < .1$, ** $p < .05$, and *** $p < .01$.
In the stats columns: * $p > .05$ and ** $p > .1$ for the null-hypothesis that $\beta_l + \beta_k = 1$.
Source: authors' calculations based on CIT-IRP5 version 4.0 (National Treasury and UNU-WIDER 2021) as discussed in (Ebrahim et al. 2021).

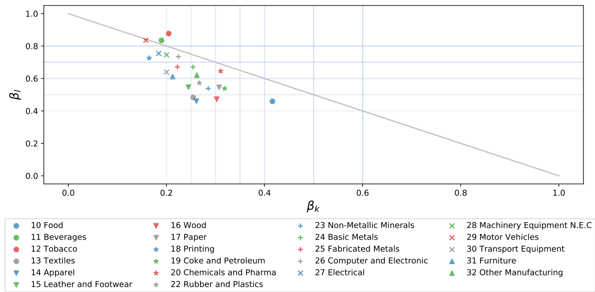
D Coefficients for other capital stock measures

Figure A5: Coefficients by estimator and type of depreciation for perpetual inventory PPE

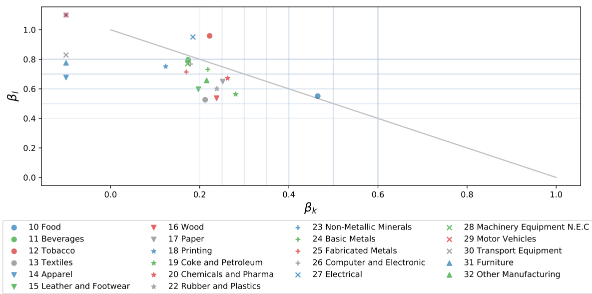
(a) OLS with capital and δ imp.



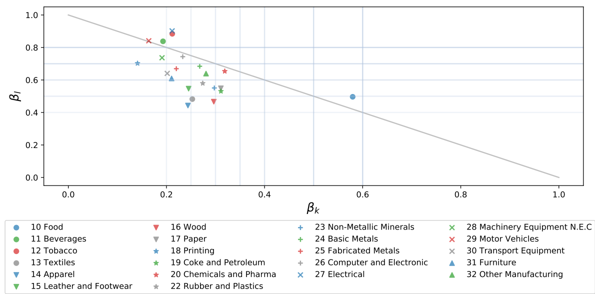
(b) OLS with capital imp. and $\delta = 10\%$



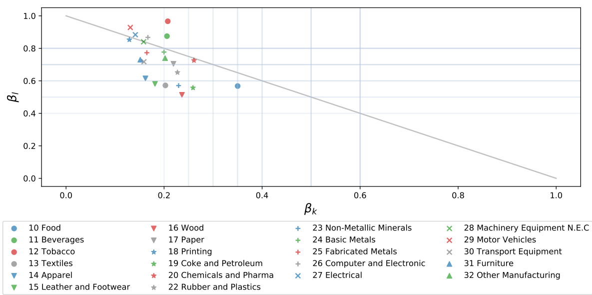
(c) PRODEST with capital and δ imp.



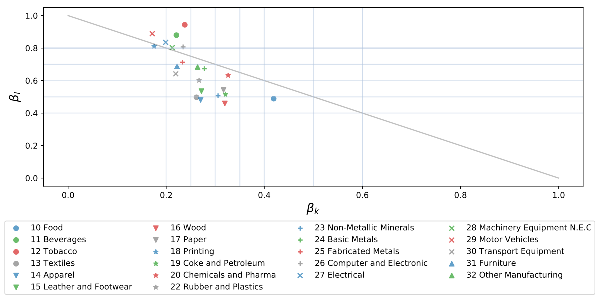
(d) PRODEST with capital imp. and $\delta = 10\%$



(e) ACF(ρ) with capital and δ imp.



(f) ACF(ρ) with capital imp. and $\delta = 10\%$



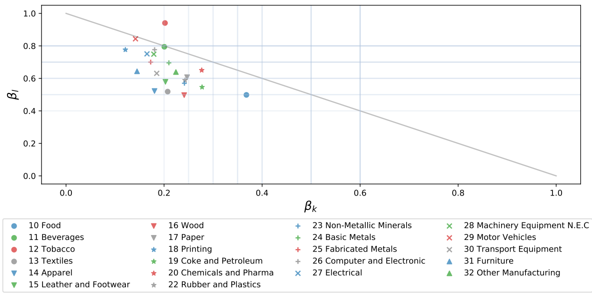
Note: this figure shows the coefficients capital, β_k, and labour, β_l, at the the 2 digit ISIC4 industry classification for OLS, PRODEST, and ACF(ρ). To simplify presentation we set $\beta > 1$ to $\beta = 1.25$ and $\beta < 0$ to $\beta = -.25$. Points to the right of the grey diagonal represent industries with increasing returns to scale, whereas lines to the left of the diagonal represent decreasing returns to scale.

The sample removes any firm that is observed to ever have a output-to-labour, output-to-capital, or capital-to-labour ratio above the 99th percentile or below the 1st percentile within the industry.

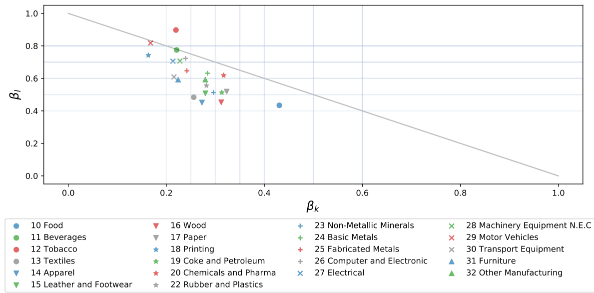
Source: authors' calculations based on CIT-IRP5 version 4.0 (National Treasury and UNU-WIDER 2021) as discussed in (Ebrahim et al. 2021).

Figure A6: Coefficients by estimator and type of depreciation for perpetual inventory fixed assets

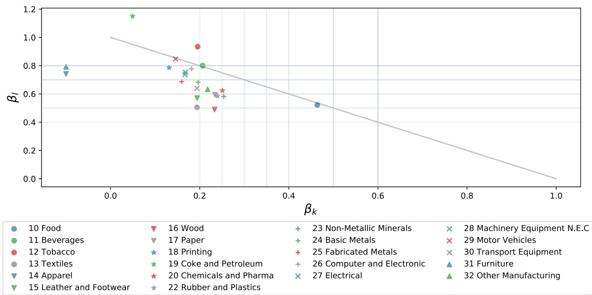
(a) OLS with capital and δ imp.



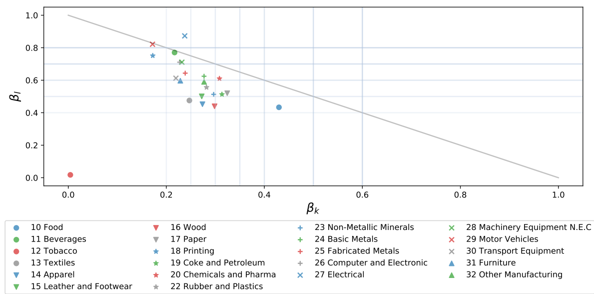
(b) OLS with capital imp. and $\delta = 10\%$



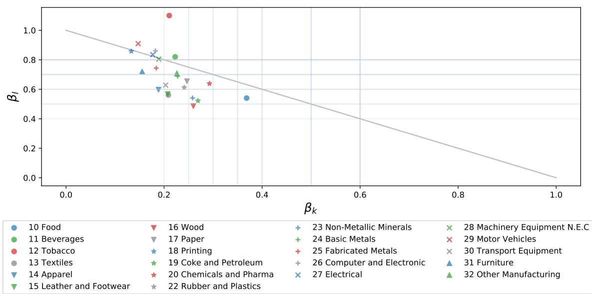
(c) PRODEST with capital and δ imp.



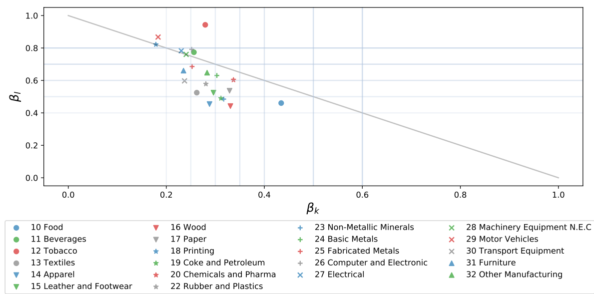
(d) PRODEST with capital imp. and $\delta = 10\%$



(e) ACF(ρ) with capital and δ imp.



(f) ACF(ρ) with capital imp. and $\delta = 10\%$



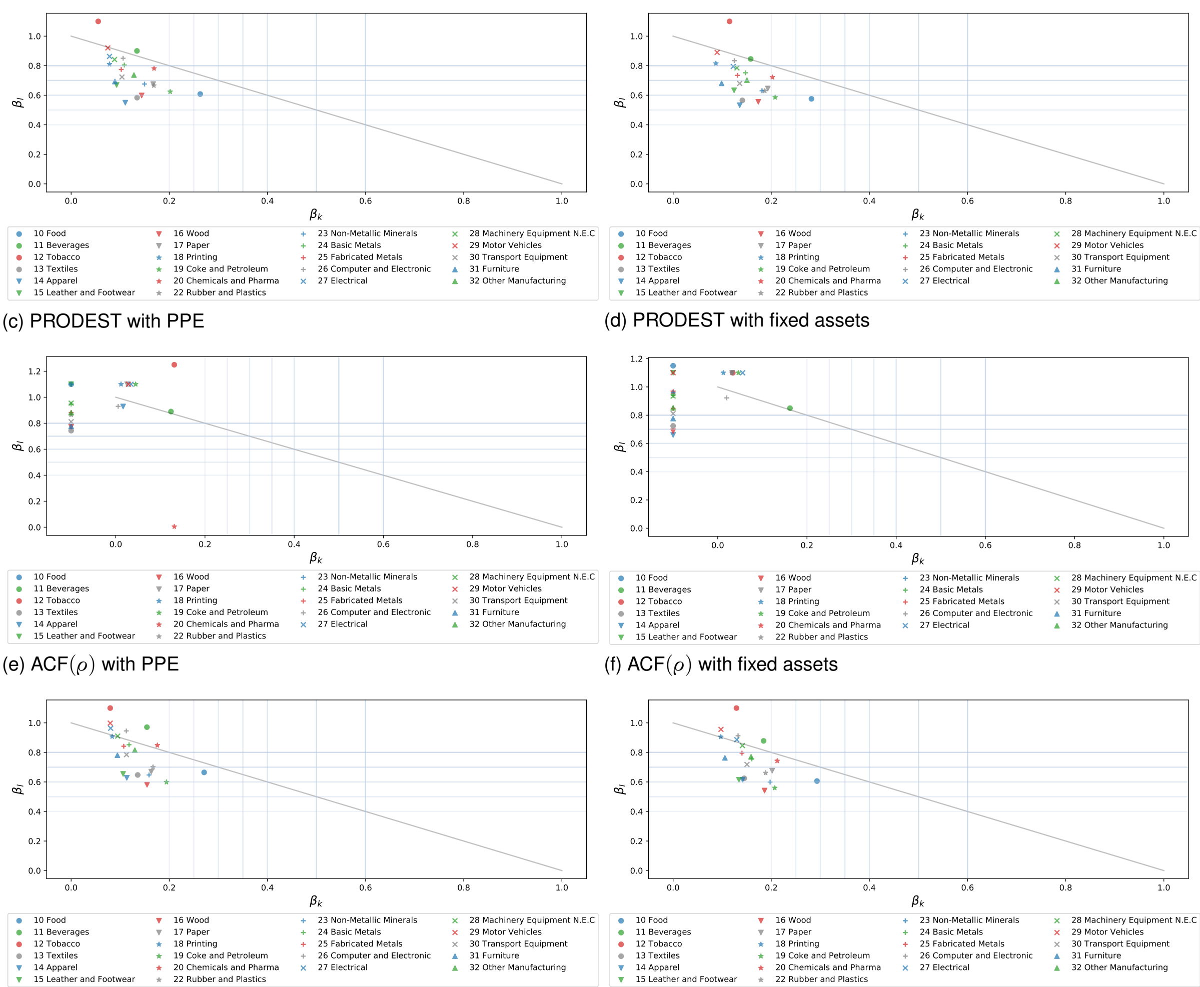
Note: this figure shows the coefficients capital, β_k, and labour, β_l, at the the 2 digit ISIC4 industry classification for OLS, on the left, and ACF, on the right. To simplify presentation we set $\beta > 1$ to $\beta = 1.25$ and $\beta < 0$ to $\beta = -.25$. Points to the right of the grey diagonal represent industries with increasing returns to scale, whereas lines to the left of the diagonal represent decreasing returns to scale.

The sample removes any firm that is observed to ever have a output-to-labour, output-to-capital, or capital-to-labour ratio above the 99th percentile or below the 1st percentile within the industry.

Source: authors' calculations based on CIT-IRP5 version 4.0 (National Treasury and UNU-WIDER 2021) as discussed in (Ebrahim et al. 2021).

Figure A7: Coefficients by estimator for fixed assets and PPE

(a) OLS with PPE (b) OLS with fixed assets (c) PRODEST with PPE (d) PRODEST with fixed assets (e) ACF(ρ) with PPE (f) ACF(ρ) with fixed assets



Note: this figure shows the coefficients capital, β_k, and labour, β_l, at the the 2 digit ISIC4 industry classification for OLS, on the left, and ACF, on the right. To simplify presentation we set $\beta > 1$ to $\beta = 1.25$ and $\beta < 0$ to $\beta = -.25$. Points to the right of the grey diagonal represent industries with increasing returns to scale, whereas lines to the left of the diagonal represent decreasing returns to scale.

The sample removes any firm that is observed to ever have a output-to-labour, output-to-capital, or capital-to-labour ratio above the 99th percentile or below the 1st percentile within the industry.

Source: authors' calculations based on CIT-IRP5 version 4.0 (National Treasury and UNU-WIDER 2021) as discussed in (Ebrahim et al. 2021).