THE GLA WORKING PAPER SERIES

Markups and Productivity.do

Do Markups Drive Productivity Under Imperfect Competition? Evidence from Eswatini

Samuel V. Mhlanga

Working Paper 00004

http://www.growthlab.uneswa.ac.sz/

THE GROWTH LAB FOR AFRICA
University of Eswatini
Kwaluseni Campus
December 2023

I thank the GLA Technical Team; namely, B. Dlamini, C. Fowler, J. Kule, A. Metfula, and L. Olarinder for invaluable comments and suggestions on an earlier draft. Any errors and conclusions expressed herein are mine and do not necessarily represent the views of the University of Eswatini or the Central Statistics Office (CSO) of Eswatini.

The GLA working papers and preprints are circulated for discussion and comment purposes. They have not undergone peer-review nor necessarily been subject to review by the GLA Editorial Board that accompanies official GLA publications.

© 2023 by Samuel V. Mhlanga. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Do Markups Drive Productivity under Imperfect Competition? Evidence from Eswatini

By Samuel V. Mhlanga
First Version: 17 December 2023

Abstract

The primary goal of this paper is to determine the impact of industrial markups on production efficiency. It relies on an optimal high-dimensional sparse modelling (HDSM) method for estimation and inference using an unbalanced dataset for the manufacturing sector supplied by the Central Statistical Office (CSO) of Eswatini. First, it found that markups and productivity declined systematically in the first four years of trade reforms and stabilized thereafter, while both objects exhibited distributional truncation from below due to the exit of unproductive plants. Second, the impact of markups (γ) on productivity was significantly positive for all plant-types, except for plants employing less than 60 workers. That is; $\gamma \in [0.00, 0.77]$ and robust to measurement choices of production efficiency and firm-type. The structural effects of markups on productivity for downsizers and large producers are circa three-times larger than the impact of markups charged by upsizing firms. In essence, downsizers became lean and meaner than expanding plants in order to survive the emerging competition within the Southern African Customs Union (SACU).

JEL Classification: D21, D24, L11

Keywords: Markups, Productivity, Technical Efficiency, High-Dimensional Sparse Modelling, Eswatini.

1. Introduction

The purpose of this article is to estimate and conduct inference on causal effects of markups on productivity shocks. At the heart of the study, the fundamental question is: Do firms invest in internal production efficiency in order to charge higher markups in future? The short answer is: Yes. Markup elasticities are significantly positive, with orders of magnitude dependant on firm-type controls.[1]

A large and growing literature beginning with Hall (1988), refined by De Loecker and Warzynski (2012) and De Loecker et al. (2020), focuses on markup measurement in the absence of information about the physical product quantity or product prices. Its basic emphasis is that firms optimize the demand function of the flexible input while treating irreversible fixed capital stock as a state variable with prohibitively high adjustment costs. The resulting first-order conditions (FOCs) from the optimization problem yield a shadow price equal to the marginal cost of production, and the measured markup is expressed as a ratio of the output elasticity for the flexible input to that input's expenditure share of total revenue. In their rebuttal of the production approach to markup determination, Bond et al. (2021) call this measure the ratio estimator.

At the same time, the estimation of production functions is not without contestations. Econometric concerns arise when technology used in production is responsive to output determinants that are unobserved to the econometrician but observed to the producer in a cost-minimization environment. One such issue is the endogeneity problem which obviates the use of ordinary squares (OLS) methods that generate biased coefficients of inputs. Related techniques on the simultaneity bias have been proposed since the last quarter century by Olley and Pakes (1996) (OP) and Levinsohn and Petrin (2003) (LP) and Ackerberg, Caves, and Frazer (2015) (ACF) and numerous others, see survey by Ackerberg, Benkard, Berry, and Pakes (2007). More recently, Rovigatti and Mollisi (2018) build on Wooldridge (2009) using dynamic panel instruments to develop an alternative proxy method while Hu et al. (2023) propose a method that is robust to optimization and measurement errors and also robust to functional dependence problems.

[1] The importance of markup studies is embodied in the status of antitrust statutes in advanced societies where the statutes are described as the "Magna Carter of free enterprise and are as important to the preservation of economic freedoms and the free enterprise system as the Bill of Rights is to the protection of fundamental personal freedoms", Baker (2003).

At the core of the analysis is that the data generating process for markups and productivity occurs under monopolistic competition, and firms facing this market structure have no supply functions (Mrazova and Neary, 2017, 2021; Foster et al.; 2018).[2] One outcome of this is that exogenous supply shifts or heterogeneous productivity shocks typically lead to elusive patterns of firm behaviour; *albeit*, with attendant implications that depend on the curvature and slope of the demand architecture. For example, if the functional form of consumer utility is the Hyperbolic Absolute Risk Aversion (HARA) with variable marginal costs, variable markups and variable output prices which, according to Perets and Yashiv (2015), is not only tractable but also an essential restriction of economic optimization. Its general structure is presented in Dhingra and Morrow (2019), and applied in Foster et al. (2018) and Mhlanga (2023). In this economic environment, a change in production efficiency may follow from price variation and shifts in the demand function.

Our empirical strategy relies on the standard production approach to measure the granular markups and use Hu et al. (2023) to measure productivity shocks. The estimation of structural effects of markups on productivity then relies on the optimal high-dimensional modelling (HDM) method with covariates potentially larger than the sample size (Belloni et al., 2016). The high-dimensional feature of the model arises from the large number of time-varying characteristics of the unit of analysis as well as higher-order series terms and their interactions as in Chen (2007), Chen and Pouzo (2012), and Newey (1997). To avoid related model overfitting problems, we gather only the non-zero regressors through dimension reduction and regularization processes to produce a parsimonious regression model that has approximate sparsity properties (Belloni et al., 2016; Belloni et al., 2012). To further eliminate the impact of correlations between observations within the same cross-sectional unit of observation among producers, we use a Cluster-Lasso estimator that accommodates the clustered covariance structure and allows for partialling out individual-specific heterogeneity (Belloni et al., 2016). We then estimate the optimal high-dimensional sparse model (HDSM) using the standard OLS as in Belloni et al. (2014) for the partially linear model or instrumental variable (IV) estimator as in Belloni et al. (2012), respectively.

This paper shares a common thread with De Loecker and Warzynski (2012) and De Loecker et al. (2020) who develop the production approach to markup estimation. It is empirically

[2] This is in contrast to perfectly competitive market environments where a shift in the supply curve translates to movement along a demand curve with the first order condition that depends on the demand elasticity. Under monopolistic competition, exogenous supply shocks lead to movements along a revenue curve and the implications of this depend on the demand manifold (Mhlanga, 2023).

connected to the proxy approach for the estimation of the Solow-residual from a Cobb-Douglass function. In determining structural effects of markups on total factor productivity (TFP), it departs from the standard IV methods which depend exclusively on economic intuition. Instead, it relates strongly to the HDSM as applied to panel data analyses in the presence of individual-specific unobserved heterogeneity (Belloni et al., 2016; Belloni et al., 2012; Belloni et al., 2014). The choice of the HDSM method is predicated on its ability to nest other approaches to parametric estimation and inference.

A preview of our findings indicates that the distribution of markups fell in the first four years and then stagnated thereafter while productivity fell persistently throughout the trade reform period. Although both variables were trimmed from the bottom due to tougher competition, they were highly skewed to the right to reflect the Darwinian effects of firm entry and exit induced by trade reforms. Similarly, the impact of markups on productivity was significantly positive and robust to measurement choices of production efficiency. These results were robust even after controlling for firm-size and firm-size variation. For instance, the structural effects of markups on productivity for downsizers and large producers were circa three-times larger than the impact of markups charged by upsizing firms. Thus, downsizers became lean and meaner than expanding plants in order to survive the emerging competition within the Southern African Customs Union (SACU).

Our contribution involves the use of a robust HDSM technique that nests other methods in firm-level panel data analysis. The flexibility of this approach enables one to estimate the HDM and select IVs/covariates using a post-double-Lasso device, estimate the IV canonical model, and test for the endogeneity hypothesis concerning the variable of interest. If hypothesis test results are affirmative, one estimates the model with Two-Stage-Least Squares (TSLS) and carry out the necessary inference. Otherwise, one estimates the canonical partial linear regression model using OLS and conduct statistical inference.

The remainder of this paper is organized as follows. Section 2 lay out the background to the economics of complexity in Eswatini. In particular, economic complexity, product complexity, and the complexity outlook indices are the focus of explanation and discussion. Section 3 discusses the theoretical motivation and econometric framework to give context to subsequent objects of analysis. Section 4 provides the framework of estimation and inference concerning the causal effects of markups on productivity. Section 5 presents the results while section 6 summarizes and concludes the paper.

2. Background to the Economics of Complexity in Eswatini

To provide context to the analysis of endogenous market structure and firm-heterogeneity in terms of idiosyncratic productivity shocks as well as their structural relationship, this section explores the country's economics of complexity during the *de facto* trade liberalization period. Our quantitative indicators for the characterization of the economy come from the Harvard Growth Lab. The fundamental thesis is that industry productivity growth and growth in Gross Domestic Product (GDP) respond positively and significantly to product upgrades, and to an increasingly more sophisticated export mix (Hidalgo et al., 2007; Mayer et al., 2021).[3] It is therefore instructive to consider at least three metrics that depict certain dimensions of economic performance; namely, Economic Complexity Index (ECI), Product Complexity Index (PCI) and Complexity Outlook Index (COI).

The ECI captures the diversification and scientific complexity of products embodied in the export basket. It therefore aligns with the notion that countries featuring diverse, complex and specialized productive knowhow tend to export diverse, complex and unique products. Perhaps the most important and arguably crucial feature for the purpose of this article is that such properties present this metric as a predictive indicator of income levels and dynamics of long-run economic growth. Its actual computation is outlined in the Harvard's Atlas of Prosperity and delivers $ECI \in [0.3,\ 0.4]$ for the HS 1992 code during the 2000-2003.[4]

Secondly, the product-specific measure known as PCI ranks economies on two dimensions. (1) As a country-specific measure, it reflects the number of products for which the country has a comparative advantage (*aka* diversity). (2) As a product-specific measure, it reflects the number of countries that have comparative advantage in the productive knowledge required for production of the product (*aka* ubiquity). The computation of this index is influenced by the number and complexity of the countries that can produce a particular product. Our PCI therefore captures the scope and complexity of knowhow required to produce the product in question. Suppose then that the upper and lower bounds of PCI are respectively presented as PCI^U and PCI_L. Then the data-driven bounds of PCI are $PCI^U \in [2.3,\ 2.6]$ and $PCI_L \in [-3.0,\ -2.7]$, $\forall\, t\ \in (2000, 2003)$. A similar distribution of PCI yields both positive and

[3] Economic growth ($\Delta lnGDP_t$) varies robustly with continuous accumulation of knowhow to help an economy diversify its production into more sophisticated activities. This arises particularly through upgrading that involves learning-by-doing/watching, product quality improvement, innovation in production processes, and adoption of suitable business technologies with possibilities of technological diffusion within and across firms.

[4] For purposes of benchmarking, the best performing economy during the same period was Japan, with $ECI \in [2.68,\ 2.82]$.

negative numbers when extending this analysis to 2016-2021, even though this lies outside our sample period. What is immediately obvious is that constrained export growth is associated with *low* PCI of major export products; e.g., Sugar and Candy, Precious Metals and Stones, as well as Apparel. Moreover, these products are traded under non-reciprocal market access conditions with pre-determined prices in the EU and US.

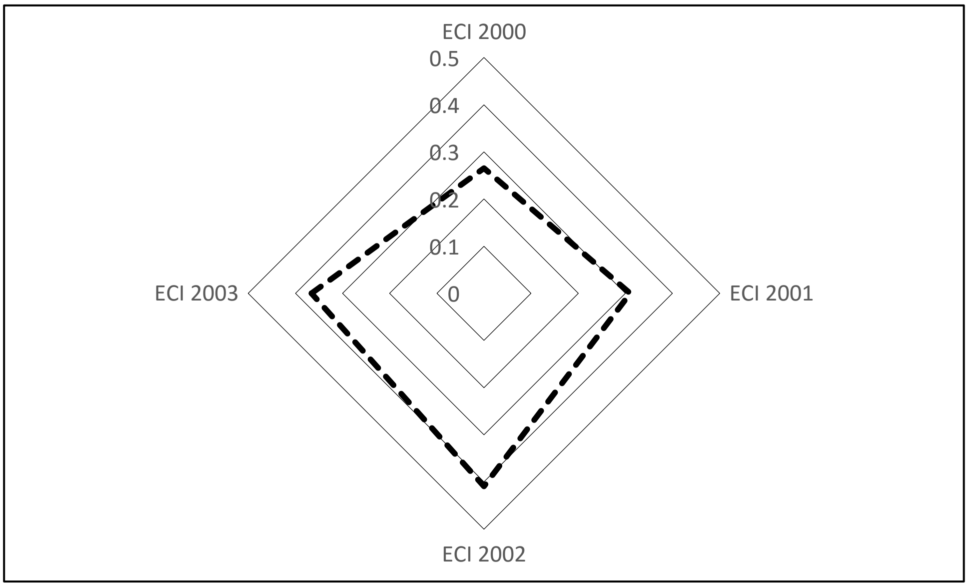

Figure 1: ECI Covering the Period 2000-2003

Source: Author's Elaboration from the Atlas of Economic Prosperity

The overall position concerning diversification into higher complexity products is reflected in the measure of the Complexity Outlook Index (COI). On the one hand, a high COI reflects the presence of capacity in the productive knowhow to produce high complexity products. On the other hand, a low COI reveals constrained capacity in productive knowhow to produce high complexity products. The null set $(COI^{+} \in [\emptyset])$ represents a measure of low product sophistication during trade liberalization, which reflects a difficulty when attempting to venture into the production of complex products.

In summary, a consolidated message from the three complexity indices (ECI, PCI, and COI) is that this economy has limited productive capability and knowhow to design and produce additional, higher complexity products. The primary nature of the dominant export commodities introduces an economic landscape of inflexibility concerning foreign demand, export volumes and price competition. This imposes the requirement to broaden the diversity and deepen the complexity of its productive capabilities and tacit knowhow in order to leapfrog into more complex products. This insight is validated by the value of the productivity level of the national export basket $EXPY_{c,t} = 0.5$; *albeit*, in the context of an inflated order of magnitude induced by superior performance of a single export product.[5] A thorough investigation identifying short-term to long-term options for product diversification and increased complexity is required.

In spite of the economic landscape just outlined, firms still consider investment in production efficiency in order to sell their product varieties at higher markups over marginal costs. The rest of the paper investigates the extent to which pricing power decisions are sufficiently flexible in the presence of a low complexity outlook of the economy.

3. Theoretical Motivation

This section examines some of the most pervasive objects of economic analysis in industrial organization; namely, markups and productivity with endogenous market structure under monopolistic competition (Berry and Compiani, 2021). In particular, it derives a production-based markup expression in the next subsection as well as derives expressions for the assumed market demand structure and revenue function in the subsequent subsection.

3.1. Derivation of the Expression for Markups

We now consider the supply- and demand-sides of the market environment to discuss the theoretical foundations for recovering markups from a production function in the absence of suitable data for calculating the ratio of firm-specific price over marginal cost and incorporate preference technology in the analysis, respectively. On the production side, DLW develop a quantitative framework for characterising markup equations by assuming that a firm's objective function entails production cost minimization. The exposition begins with a firm-specific output index over time, Q_{it}, that is continuous and twice differentiable, a static freely variable production input X_{it}^{v}, a dynamic stock of state capital K_{it}, and a Hicks-neutral idiosyncratic productivity shock to multiproduct firms, ω_{it}. However; in the case of single product firms, the Hicks-neutral condition is redundant (see De Loecker *et al.* (2016)). Setting up the Lagrangian

[5] The product-level measure of sophistication expressed as GDP_t per capita weighted by revealed comparative advantage (RCA) was developed by Hausmann, Hwang, and Rodrik (2007) as $PRODY_{i,t} = \sum_c \left[\left(xval_{c,i,t} / \sum_i xval_{c,i,t} \right) / \sum_c \left(xval_{c,i,t} / \sum_i xval_{c,i,t} \right) \times Y_{c,t} \right]$, where $xval_{c,i,t}$ = product i for country c at year t, and $Y_{c,t} = GDP_t$ per capita for country c at year t. The authors evaluated the income-productivity level associated with the export basket as $EXPY_{c,t} = \sum_i \left(xval_{c,i,t} / \sum_i xval_{c,i,t} \right) PRODY_{i,t}$ for Eswatini. As it turns out, the size of this measure was influenced by a top performing product "*Mixed Odoriferous Substances in the Food and Drink Industries*" with a high value of $PRODY_{i,t}$.

function $\left(\mathcal{L}_{it}(\cdot)\right)$ for a firm's cost-minimization problem produces a first-order condition (FOC) with respect to the freely adjustable input to yield three results. First, it equates the price of the freely adjustable input $\left(P_{it}^{X^v}\right)$ to the marginal product of X_{it}^v *times* the Lagrange multiplier $\left(\frac{\partial \mathcal{L}_{it}(\cdot)}{\partial Q_{it}} = \lambda_{it}\right)$, where $\frac{\partial Q_{it}(\cdot)}{\partial X_{it}^v}$ is firm $i's$ marginal product of input X_{it}^v at time t, and λ_{it} is the standard shadow price representing the marginal cost of production at a given firm's output. Second, one rearranges terms and multiplies on both sides by $\frac{X_{it}^v}{Q_{it}}$, conditions on state variables, dynamic capital stock, then the output elasticity of the freely variable input $\theta_{it}^{X^V}$ becomes

$$\theta_{it}^{X^V} = \lambda_{it} \frac{P_{it}^X X_{it}}{P_{it} Q_{it}}$$

Third, the firm-level markup at time t is defined as $\mu_{it} = \frac{P_{it}}{\lambda_{it}}$, which can therefore be measured as

$$\mu_{it} = \frac{\theta_{it}^X}{\alpha_{it}^X}$$

where α_{it}^X is the ratio of freely variable input expenditure to sales revenue.

In order to estimate markups using intermediate inputs, one relaxes the assumption of fixed proportionality production technology of material inputs to output. As in De Loecker *et al.* (2016), this seems sensible given the level of aggregation of the dataset at hand which allows for substitution of labour for capital while keeping output unchanged. Such a specification of functional form involves estimating a gross output production function by using multiple FOCs to recover markups $\hat{\mu}_{it}^m$ through output elasticity of intermediate inputs as

$$\hat{\mu}_{it}^m = \hat{\beta}_m \left(\frac{P_{it}^m M_{it}}{P_{it} Q_{it}}\right)^{-1}$$

where P_{it}^m refers to the unit price of material (M_{it}) cost for firm i at time t. Our preference for material inputs as a freely variable input derives from the expectation of a higher frequency of material procurement for production compared to churning in the hiring and termination of labour services. Furthermore, the gross production function requires that material input prices vary across firms and are serially correlated over time. Lazear (1990) posits a strict condition that requires perfect functioning of markets such that anything less remains distortionary in the goods markets.

Markup estimates based on the production approach are susceptible to the direction of change in the expenditure patterns of freely variable inputs as well as in output elasticities with respect to such inputs. For instance, markup computations based on intermediate inputs are a decreasing function of intermediate expenditure shares of revenue, but an increasing function of the output elasticity with respect to these inputs. That is, $\uparrow \hat{\mu}_{it}^{m} = g\left(\uparrow \hat{\beta}_m, \downarrow \frac{P_{it}^{m} M_{it}}{P_{it} Q_{it}}\right)$, in short form. The argument in $g(\cdot)$ can move simultaneously to exert an even stronger effect on markups. As noted by Basu (2019), the downside to these properties is that they have the effect of producing implausible orders of magnitude in markups over time. In datasets where the marginal cost is invariable to product quantity fluctuations, the output elasticity is also valued at the cost-share of input (De Loecker and Warzynski, 2012; Foster et al., 2017).

An alternative measure of price over marginal cost for every plant at any point in time in occasions of data sparsity is related to the Lerner index. This method is particularly potent if the incidence of negative, zero, and inestimable output elasticities is high. It is therefore worthwhile to understand its relationship, if any, with the production-based method of markup estimation. Mhlanga and Rankin (2021) demonstrate that $\mu_{it}^{pcm} = \frac{P_{it} - C_{it}}{C_{it}} = \hat{\mu}_{it}^{m}$ is true if, and only if, $\left(\hat{\beta}_m = \frac{P_{it} - C_{it}}{P_{it}}\right)$. The Lerner index becomes important in empirical environments featuring non-negligible elements of data sparsity because it is convertible to observed revenues and variable costs. However, this article focusses on measurement and analysis of $\hat{\mu}_{it}^{m}$, unless stated otherwise.

3.2. Consumer Demand and Production under Monopolistic Competition

On the demand side, we introduce monopolistic competition with heterogeneous markups to the relevant markets populated by differentiated product varieties and consumers whose preference profile fits the HARA proposition.[6] More specifically, we work with the general form of the HARA utility function to characterise its inverse demand function

$$P_{it} = \xi_{it} \times \left(\frac{Q_{it}}{1 - \rho_{it}} + \alpha_{it}\right)^{\rho_{it} - 1}$$

[6] This is distinct from Hsieh and Klenow (2009) who assume isoelastic inverse demand technologies with negative unit elasticity (-1) with respect to technical efficiency; i.e., $TFPQ_i$, flat marginal costs, and production costs and demand structures that are uncorrelated with distortions $(TFPR_i)$, for each firm i.

where a plant-specific demand shifter is ξ_{it}, $\alpha_{it} > \frac{lnQ_{it}}{\rho_{it}-1}$, $\rho_{it}\epsilon(0,1)$, and $\varepsilon_{it}^{D} = \rho_{it} - 1$ are time-variant parameters.[7]

From the underlying consumer utility structure, the theoretical foundations of Peters (2020) for systematic idiosyncratic heterogeneity in markups are accommodated. Thus, profit maximization under monopolistic competition produces micro variability of markups as

$$1 + \mu_{it} = \frac{Q_i + \alpha_{it}(1 - \rho_{it})}{\rho_{it}Q_i + \alpha_{it}(1 - \rho_{it})}$$

and is not invariant to output fluctuations, perturbations in the price elasticity of demand, and variations in the shape of the demand curve. Solving the expression above, the individual objects ihe Markup Equation for the triplet $(\rho_{it}, \alpha_{it}, Q_{it})$ produces:

Demand Parameter (ρ_{it}): $\rho_{it} = g_\rho(\alpha_{it}, Q_{it}, \mu_{it}) = \frac{Q_{it}-\alpha_{it}\mu_{it}}{Q_{it}-\alpha_{it}\mu_{it}+Q_i\mu_{it}}$,

Shape Parameter (α_{it}): $\alpha_{it} = g_\alpha(\rho_{it}, Q_{it}, \mu_{it}) = \frac{Q_{it}-\rho_{it}Q_{it}(1-\mu_{it})}{\mu_{it}(1-\rho_{it})}$

and

Physical Quantity of Output (Q_{it}): $Q_{it} = g_q(\alpha_{it}, \rho_{it}, \mu_{it}) = \frac{\alpha_{it}\mu_{it}(1+\rho_{it})}{1-\rho_{it}}$

Thus, the Markup Equation is an increasing function of demand shocks and production technologies embodied in output for all firms and products under HARA preferences. As noted by Syverson (2019), the size of markups at profit maximizing outputs is determined by the shape of the inverse demand (α_{it}) that firms face, the demand elasticity (ε_{it}^{D}), and quantity of output demanded (Q_{it}). If the demand elasticity changes with output when plant heterogeneity in productivity characterizes the industry, then this creates variation in the pricing power of firms (Dhingra and Morrow, 2019). Moreover, the steeper the inverse demand curve, the higher the pricing ability is for the firm facing that demand structure. In the enquiry by Mrázová and Neary (2017), the smooth curve that relates the demand elasticity and the shape of the demand structure; *aka* the demand manifold $(\varepsilon_{it}^{D}, \alpha_{it})$, is a sufficient statistic for a significant segment of comparative statics predictions. Moreover, Beggs (2021) demonstrates that the demand manifold

7 Perets and Yashiv (2015) assert that a HARA demand structure is not only tractable, but it is a fundamental economic necessity.

and related comparative statics predictions are strictly invariant to structural demand variations on condition these adjustments arise from changes in market size and changes in product quality.

The ideal characterization of the HARA economy is predicated on the full flexibility of the Demand Manifold to allow for unconstrained monopolistic competition as Mhlanga (2023) demonstrates. However, our information infrastructure is less than ideal. The demand-side computational requirement is that the functions $g_{\alpha}(\cdot)$ and $g_{\rho}(\cdot)$ are fully known. Yet, this is not the case. In particular, physical output Q_{it} is an empirical object that is generally hard to observe in datasets. In this sense, we follow Foster et al. (2017) and proxy it with deflated industry output to resolve the unobservability problem.

Furthermore, we fix the markup at $1 + \mu_{it} = 1 + \hat{\mu}_{it}^{m}$, where $\hat{\mu}_{it}^{m}$ is the DLW estimate at the firm-level. To then compute the Demand Parameter that determines ε_{it}^{D}, one needs to know the distributional patterns of the Shape Parameter, α_{it}. Given the narrow support of the Shape Parameter as clearly shown by the lower and upper bounds of the elements of the empirical demand manifold, it seems sensible to fix the parameter at its first-order moments. More specifically, one assigns the median value to collect all the ingredients necessary for the computation of the Demand Parameter used to scale revenue elasticities and the measured productivity based on the revenue function. Fixing α_{it} also fixes the demand elasticity. This is not a serious caveat given the range of *median* inverse elasticities of demand for various firm-types reported in Mhlanga (2023).

On the production side, the measure of technical efficiency is $TFPQ = Q_{it} - Af(A_{ik}K, A_{il}L)$, where A is Hicks-neutral technology, A_{ik} and A_{il} are respective capital- and labour-augmenting technologies. For simplicity, we assume $A_{ik} = A_{il} = 1$.[8] In tandem with the literature, we assume isoelastic demand and non-constant marginal costs (MCs) so that a residual demand shift leads to a commensurate shift in product quantity and product price. This pattern of MC movement implies that the revenue-based multifactor productivity (TFPR) varies too, with the correlation structure only dependent on whether MC is increasing or declining. The implicit

[8] Demirer (2022) postulates production technologies of the form $Y_{it} = F_t(K_{it}, \omega_{it}^{L} L_{it}, M_{it}) exp(\omega_{it}^{H}) exp(\epsilon_{it})$, where K_{it} and L_{it} are firm-level factor inputs, $M_{it}$is material input, $\omega_{it}^{L} \in \mathbb{R}_{+}$ is labour-augmenting technology, $\omega_{it}^{H} \in \mathbb{R}$ denotes Hicks-neutral productivity shocks and $\epsilon_{it} \in \mathbb{R}$ is the standard random shock to output, Y_{it}. Although the specification of Y_{it} excludes capital-augmenting technology $(\omega_{it}^{K} \in \mathbb{R}_{+})$, which might be important in economic settings with robust Foreign Direct Investment (FDI) inflows, it still nests many of the production functions used in the literature.

value-added Cobb-Douglas revenue function for this economy is $P_iQ_i = P_sQ_s^{1-\rho}Q_i^{\rho-1}\xi_iQ_i = P_sQ_s^{1-\rho}Q_i^{\rho}\xi_i$. In logs, this translates to

$$p_i + q_i = p_s + (1-\rho_i)q_s + \rho q_i + ln\,\xi_i = p_s + (1-\rho)q_s + \rho\left(\sum_{r=1}^{R}\alpha_{ir}^{v}x_{ir} + a_i\right) + ln\,\xi_i$$

where r refers to sources of confounding variation on revenue. Typically, the literature estimates a revenue function with parameter estimates β_r for each input r and estimates the regression-based total factor productivity (TFP^{rr}); see, for example, Levinsohn and Petrin (2003). As is obvious from this expression, after purging the firm-level subscript i, $\rho\alpha_r^v = \beta_r$, implying $\alpha_r^v = \beta_r/\rho$ and $TFP^{rr} = \rho \times TFPQ$, indicating that β_r embodies factor elasticities and confounding demand parameters. As argued by Foster et al. (2016), TFP^{rr} is indeed a function of fundamentals $(TFPQ)$ and idiosyncratic demand shocks such as ρ, which is the parameter that determines the member of the demand manifold ε_{it}^{D} faced by the firm.

Another valuable representation of productivity is the total factor productivity (TFP^{cs}) based on cost-shares of factor input expenditures under cost-minimization objectives and constant returns to scale. This measurement approach to production efficiency relates more strongly to cross-plant averages and long-term patterns than to short-run volatility. As in Syverson (2011) and Foster et al. (2017) we make the additional assumption that FOCs of the cost-minimization problem hold up to the first-order moments; i.e., the mean, instead of expecting FOCs to hold at the granular level and every point in time. Incidentally, although we do not conduct this analysis, but these assumptions are robust to the method of econometric estimation and to the measure of production efficiency (Foster et al., 2017). There is similarity in the cross-sectional and longitudinal distributions of $TFPQ$ and TFP^{rr}, while $TFP^{rr} \neq TFP^{cs}$ also holds with certainty.

Thus, the next section focuses on inferential methods based on high-dimensional panel data modelling, outlines data sources and measurement choices, and deploys the econometric methods to uncover the structural effects of the object of interest.

4. Framework for Causal Effects of Markups on Productivity

The estimating equation adopted for this analysis that allows for markup endogeneity is

$$lnTFPQ_{it} = \alpha_0 + \beta_0\mu_{it-1} + \boldsymbol{X_{it}} + e_i + \varepsilon_{it}, \forall\, t \in T$$

where α_0 and β_0 are parameters to be estimated, e_i represents time-invariant firm-specific fixed effects, and ε_{it} denotes random noise. The key variables are technical efficiency and markups, $(lnTFPQ_{it}, \mu_{it-1})$, where the parameter of interest is β_0. The empirical model also admits a vector of controls/instrumental variables (IVs), $\boldsymbol{X_{it}}$, which enter the model according to whether exogeneity/endogeneity assumptions hold for markups, respectively. Since we do not know *ex ante* or even *a priori* what firm characteristics should enter as controls or IVs in the equation, not to mention what the nature of their interaction or transformation must be, we introduce splines, power series, and Fourier series, as well as trends, dummies, and interactions in the analysis. If there are p high-dimensional controls/IVs and a sample size n, the HDSM procedure admits potentially $p \gg n$ series terms from which regularization devices objectively select only $s \ll p$ approximately sparse series terms (Belloni et al. 2012; and Belloni, Chernozhukov, and Hansen, 2014b). Regularization in this instance entails ℓ_1-penalization of each variable in the chosen estimator, typically the Least Absolute Shrinkage and Selection operator (Lasso) that minimizes the objective function of the sum of squared residuals and a penalty term (Belloni, Chernozhukov, and Hansen, 2014b). The high-dimensional data (HDD) generated by this process renders the use of ordinary least squares (OLS) inappropriate for application in the estimating technology due to potential overfitting problems and omitted variables' bias.

Our aim is to explore the impact of markup variability on changes in idiosyncratic productivity shocks by relying on two hypotheses. First, endogenous markups are subject to optimal IVs with additive fixed effects (Belloni, Chernozhukov, Hansen, Kozbur, 2016). Second, technical efficiency is driven by directly by exogenous markups and $s \ll p$ optimally selected high-dimensional sparse controls. As Belloni, Chernozhukov, Hansen, Kozbur (2016) argue, even though the unbalanced nature of the dataset complicates the analysis, it is sensible to assume that observations are missing-at-random in order to resolve this structural data problem.

Therefore, we combine economic intuition with high-dimensional sparse modelling (HDSM) to uncover the identities of IVs and controls that act as sources of variation in endogenous markups and in unobserved idiosyncratic technical efficiency shocks, respectively. In an environment of robust market competition induced by trade reforms, incumbent firms may opt to invest in the quality dimension of their product mix in order to subsequently raise the price of product varieties at a later time (Peters, 2020). In essence, the effect of such investment policy might come in a form of induced variation in product demand and production cost fundamentals so as to move firms in profitable directions thereby increasing markups and augmenting productivity dispersion (Haltiwanger et al., 2018). In order to elicit a true relationship between product quality and price over marginal cost, one can benefit from high-dimensional constructs of series terms (Chen, 2007).

4.1. Productivity and Markup Endogeneity with Optimal IVs

The estimation of structural parameters and associated inference under markup endogeneity assumptions derives its key features primarily from Belloni, Chen, Chernozhukov, and Hansen (2012) and Belloni, Chernozhukov, Hansen, Kozbur (2016) specifically for panel data.[9] The variation in productivity shocks, ω_{it}, remain a function of $t-1$ endogenous markup variation (μ_{it-1}), changes in exogenous cofounders, evolution of additive time-invariant individual-specific heterogeneity (e_i), and random noise (ε_{it}) volatility. In the first-stage, markups respond only to a few of a plant's time-variant characteristics, time-invariant individual plant-specific heterogeneity, and random disturbances. That is, uninfluential IVs and fixed effects are partialled out at the variable selection stage.

To purge the model of fixed effects, we assume that the missing values in the unbalanced panel dataset conform to the missing-at-random hypothesis so one can demean productivity and confounders for each plant i: $ln\ddot{TFPQ}_{it} = lnTFPQ_{it} - \frac{1}{T}\sum_{t=1}^{T} lnTFPQ_{it}$ and $\ddot{\mu}_{it} = \mu_{it} - \frac{1}{T}\sum_{t=1}^{T}\mu_{it}$, respectively.[10] It is convenient to define a function $\ddot{h}(w_{it})$ as $\ddot{h}(w_{it}) := \ddot{\boldsymbol{z}}_{it}'\pi + \ddot{r}(w_{it})$, where $z_{it} = z(w_{it})$ denotes a collection of transformations of the central instrument w_{it} in the genre of Newey (1997) and Chen and Pouzo (2012). That is, w_{it} allows for the inclusion of basis functions. The data generating process (DGP) of the demeaned model therefore becomes

$$\ddot{\omega}_{it} = \gamma\ddot{\mu}_{it} + \ddot{\boldsymbol{x}}_{it}'\beta + \ddot{\boldsymbol{\varepsilon}}_{it}$$

$$\ddot{\mu}_{it} = \ddot{\boldsymbol{z}}_{it}'\pi + \ddot{r}(w_{it}) + \ddot{u}_{it}.$$

Demeaning IVs to predict demeaned markups leads Cluster-Lasso coefficients, π, to have a high probability of satisfying the necessary approximate sparsity conditions (Belloni, Chernozhukov,

---

[9] A rich and growing literature on HDSM estimation and inference already exists, with Belloni, Chen, Chernozhukov, and Hansen (2012), Belloni, Chernozhukov, Hansen, Kozbur (2016), and Belloni, Chernozhukov, and Hansen (2014b) adopting penalized estimation methods while Zhang and Zhang (2014), Javanmard and Montanari (2013), and Caner and Kock (2018) rely on *desparsification* to produce theoretical results. This article; however, chooses the panel data fixed effects modelling approach to HDMS proposed by Belloni, Chernozhukov, Hansen, Kozbur (2016) which demeans the response variable and the confounding factors prior to post-double model selection instead of using the other suitable methods such as correlated random effects suggested by Kock (2016), penalized GMM by Caner and Zhang (2014).

[10] Kock and Tang (2019) preserve the individual-specific heterogeneity in their dynamic panel data model by neither demeaning nor differencing the time-invariant fixed effects to allow for further analysis and inference on them.

Hansen, and Kozbur, 2016). Under the maintained endogeneity hypothesis of markups, and in the presence of non-zero IVs selected by Cluster-Lasso, the parameter of interest, γ, is estimable by the standard Post-Cluster-Lasso IV estimator.

4.2. Partially Linear Treatment Model of Markups and Productivity

In instances where the endogeneity hypothesis of markups is invalid, the exogenous structural model of markups and production efficiency is recast in the framework developed by Belloni, Chernozhukov, and Hansen (2014b) for time-series and Belloni, Chernozhukov, Hansen, and Kozbur (2016) for panel data. More specifically, the procedure implies a direct introduction of nuisance functions in the model; that is, $g(z_{it})$ and $m(z_{it})$ that give rise to the equation for idiosyncratic productivity $\omega_{it} = \mu_{it}\alpha + g(z_{it}) + e_i + \tau_{it}$ and the markup equation $\mu_{it} = m(z_{it}) + f_i + u_{it}$, where f_i and u_{it} denote fixed effects and the error term, respectively. Thus,

$$\ddot{\omega}_{it} = \ddot{\mu}_{it}\gamma + \ddot{\boldsymbol{x}}'_{\boldsymbol{it}}\beta + \ddot{\boldsymbol{\varepsilon}}_{it}$$

$$\ddot{\mu}_{it} = \ddot{h}(w_{it}) + \ddot{u}_{it} = \ddot{\boldsymbol{z}}'_{\boldsymbol{it}}\pi + \ddot{r}(w_{it}) + \ddot{u}_{it}$$

These equations capture the implicit assertion in Belloni et al. (2017) that the conditional expectations of demeaned productivity and markups over marginal costs, $E_P[\ddot{\omega}|\ddot{x}]$ and $E_P[\ddot{\mu}|\ddot{x}]$, are approximately sparse up to an infinitesimal approximation error, $\ddot{r}(z_{it})$. As a result, these expressions can be written in estimable form as

$$\ddot{\omega}_{it} = \ddot{\boldsymbol{x}}'_{\boldsymbol{it}}\pi + \ddot{r}(z_{it}) + \ddot{\delta}_{it} \qquad \hat{I}_{RF}$$

$$\ddot{\mu}_{it} = \ddot{\boldsymbol{x}}'_{\boldsymbol{it}}\beta_m + \ddot{r}_m(z_{it}) + \ddot{u}_{it} \qquad \hat{I}_{FS}$$

$$\hat{I} = \hat{I}_{RF} \cup \hat{I}_{FS} \cup \hat{I}_{AS}$$

where the significance of the elasticity of demand in the Demand Manifold $(\varepsilon^D_{it}, \alpha_{it})$ discussed in the previous section places its demeaned form in the amelioration set; i.e., $\ddot{\varepsilon}^D_{it} \in \hat{I}_{AS}$.[11] We then run the least squares of $\ddot{\omega}_{it}$ on $\ddot{\mu}_{it}$ and the set of controls in $\hat{I}$ as advocated by Belloni, Chernozhukov, and Hansen (2014b) and Belloni and Chernozhukov (2013). Moreover, the producer pricing schemes over marginal cost control for zero price hikes or zero markdowns, among other things.

5. The Data and Factor Input Moments

[11] Recall that the purpose of regularization is dimension reduction in order to control model overfitting. The control variables in the amelioration set should not dominate the set of controls selected by Cluster-Lasso (Belloni, Chernozhukov, and Hansen, 2014b).

This section discusses data issues and behavioural patterns of factor inputs in relation to pure profits. It also discusses and documents factor input serial correlation as well as correlation of inputs with profit rates.

5.1. The Data and Measurement Issues

The assessment of manufacturing firms' micro dynamics is based on a unique and unbalanced panel dataset drawn from the system of national accounts. This dataset is provided by the Central Statistical Office (CSO) of Eswatini and reported at the Four-Digit International Standard Industrial Classification (ISIC) covering the years 1994-2003. It provides information on the value of domestic and foreign export sales, number of paid employees and working proprietors, salaries and wages, investment flows, and expenditure on material inputs. A full description of the data is contained in Mhlanga and Rankin (2021).

Following the tradition of Haltiwanger and Cooper (2006), the measurement of investment series is

$$I_{ijt}^{k} = EXP_{ijt}^{k} - RET_{ijt}^{k}$$

where EXP_{ijt}^{k} reflects real gross expenditure on class k of the capital asset and RET_{ijt}^{k} is real gross retirements of class k the capital assets. The schedule of investments has the property that $\Delta I_{ijt}^{k} > 0$ if, and only if, $\Delta EXP_{ijt}^{k} > \Delta RET_{ijt}^{k}$ $\forall\, k \in (Land, PME, Trneq, Fureq)$.[12] At the same time, the computational dynamics of capital stock, K_{ijt}^{k}, are an outcome of the Perpetual Inventory Method (PIM) that yields

$$K_{ijt}^{k} = I_{ijt}^{k} + (1 - \delta^{k})K_{ijt-1}^{k},$$

where the object δ^{k} represents the rate of depreciation of asset class k, and K_{ijt-1}^{k} is the previous year's capital stock in industry j. In circumstances where this condition does not hold, then the firm(s) experiences either an episode of investment inaction or disinvestment. That is, $\Delta EXP_{ijt}^{k} \leq 0$ implies $I_{ijt}^{k} \leq 0$ and $K_{ijt}^{k} \downarrow$. This decline in capital inflows can also obtain when ΔRET_{ijt}^{k} rises faster than ΔEXP_{ijt}^{k}. However, the property that capital assets are industry-specific and therefore irreversible due to high capital adjustment costs, constrains their flexibility when disinvestment decisions arise. As a result, capital retirement becomes more costly relative to

12 The acronyms PME, Trneq and Fureq represent Plant, Machinery, and Equipment; Transport Equipment, and Furniture and Equipment, respectively.

capital procurement. Thus, one can focus on moments and serial correlation of factor inputs as well as their relationship with profit shocks to understand patterns of input markets.

5.2. Factor Inputs, Profits and Related Correlations

In the process of determining the relationship between multifactor productivity and markups, the need for understanding the distributional behaviour of factor-inputs in relation to production efficiency is inevitable. For instance, a key characterization of factor inputs deduced from proxy methods of estimating production technologies is that labour is quasi-fixed while land and capital stocks are fixed inputs. These properties are informative about the extent of inertia in the variables and provide a channel for the propagation of intertemporal shocks. Thus, fixed inputs tend to feature stronger serial correlation.

Table 1 reports statistics on investment expenditure, capital retirement and the stock of labour input. As shown in the PIM of capital stock accumulation, the direction of change in productive capital stock depends on plant-level investment or disinvestment in capital. The table also shows that expenditure on these capital assets is higher than the capital disposal component, with the effect of raising the overall capital stock and account for circa 37.7% of the median labour coefficient. Given this investment-labour proportionality, the manufacturing production in this economy is mostly labour-intensive. Furthermore, and consistent with firms' profit maximization objectives, there is an inverse relationship between profit rates and components of factor inputs. This reflects the effect of an increase in the procurement of factor input services on total revenue. Putting it more succinctly, an increase in factor input acquisitions raises production costs, reduces business revenues, and reduces production efficiency; hence the inverse relationship between factor input costs and profit rates.

Looking at the profitability-input relationship in the third row, there is no real correlation between capital investment/disinvestment and business profitability. To understand this, it is important to understand the measurement of profit shocks. Assume a profit function of the form: $\Pi(A_{it}, K_{it}) = A_{it}K_{it}^{\theta}$, where $a_{it} = ln(A_{it})$ and $a_{it} := b_t + \varepsilon_{it}$ denotes idiosyncratic shocks to profitability and heterogeneous labour costs. A further assumption is that the firm-specific shock ε_{it} is first-order autoregressive. After taking logs of the profit function and then taking differences, the resulting structural equation is amenable to estimation by the General Method of Moments (GMM).[13] This exercise produces the coefficients in the third row. Thus, the apparent disconnect between investment decisions and profit gains is not an assault on profit

[13] For details on this procedure, see Cooper and Haltiwanger (2006).

maximization goals achievable through capital investments. Rather, it is a reflection of special economic circumstances of weak and therefore ineffective capital inflows needed to induce firm-level profitability during trade reforms. The labour input; however, yields a significant plant-level correlation parameter with profitability thereby reinforcing the *prima facie* conclusion that the manufacturing sector was generally labour-intensive in the sample period.

Table 1: Factor Inputs, Pure Profits and Associated Correlations

Variable	Investment		Labour
	Expenditure	Retirement	
Median Value	12.10 (0.12)	10.87 (0.24)	3.26 (0.05)
Profit Rates	-0.83 (0.03)	-0.95 (0.06)	-0.86 (0.02)
Correlation with Profit Shocks	0.07 (0.14)	0.09 (0.39)	0.19*** (0.00)
Serial Correlation	-0.20*** (0.04)	-0.17 (0.12)	0.02*** (0.01)

Note: The second row refers to Pure Profit $:= (Revenue - Total\ Cost)/Revenue$, where Total Cost admits only variable costs. The correlation of profit rates with dimensions of factor inputs is measured against the levels of variables instead of their transformations. Denote investment rate as $InR_{ijt}^{k} := I_{ijt}^{k}/K_{ijt-1}^{k}$. As in the text, the presence of serial correlation in investment rates is represented by $\Delta InR_{ijt}^{k} := \varepsilon_{ijt}^{k} \neq 0$, where Δ denotes a 'change' in the variable, $\varepsilon_{ijt}^{k} = \varphi_0 + \varphi_1 \varepsilon_{ijt-1}^{k} + \zeta_{ijt}^{k} + \tau_{ij}^{k}$, $\forall \varphi_1 \in [-1,\ 1]$, τ_{ij}^{k} is individual firm fixed effect, and the orthogonality condition $\mathbb{E}\left[\zeta_{ijt}^{k} | \varepsilon_{ijt-1}^{k}\right] = 0$. Hence, $\varphi_1 = -0.57^{*}$ means there is negative serial correlation in investment flows. Robust standard errors are in brackets.

Source: Author's Calculations

Furthermore, the -0.57 serial correlation coefficient observed in capital investments is driven largely by the negative serial correlation in the expenditure component of investment capital stock. On the other hand, capital stock disinvestments exhibit an insignificant autocorrelation coefficient potentially due to high capital adjustment costs leading to sequential irreversibility of capital stock (Bertola and Caballero, 1994). In other words, industry-specificity of capital renders it very little or near-zero value once installed, unless it is used in production. The slow decay exhibited by capital investment provides further reason to demean dynamic covariates in the next section.

6. Main Empirical Results

The next subsection considers the preliminary performance of plants in sales revenue, markups, and technical efficiency under scale- and growth-dependent characteristics. Hereon, we dissect the distribution of plants and collect employers of less than 60 workers as well as large plants as defined in the revised SME Policy (2018). This generates a size composition of 69.1% SMEs. A further separation between firm contraction and expansion generates 52.8% downsizers as measured by downward employment changes.

To understand the behaviour of the distribution of covariates of interest, we present and discuss a related visual diagram. This helps answer several auxiliary questions by focusing on the present characterization of firms. In particular, do large and expanding firms dominate SMEs and downsizers in markups and production efficiency as they do *a priori*? Do SMEs face the same demand structure as downsizers, and do large firms face the same inverse elasticity of demand as upsizing plants?

Furthermore, the subsequent penultimate subsection disciplines the data with one of the fundamental questions in economics today: do markups drive variation in firm-level production efficiency and, if so, are structural effects of markups robust to other measures of production efficiency? To answer this question, we run regressions of technical efficiency and alternative measures of productivity on production-based markups using HDSM methods and a pooled dataset. We also run robustness checks for key results by regressing technical efficiency on markups under the same firm-types; i.e., splitting data by firm-size and scale-adjustment.

6.1. Exploratory Results

The industrial description as outlined in the background section has a characteristic of constrained overall position in the product space in terms of its ability to move towards more complex products; e.g., the null-set for the positive Complexity Outlook Index (COI). This subsection investigates if sector-wide and product-level complexities translate into commensurate patterns in sales revenue, markups, and idiosyncratic production efficiency. A potential example is a decline in product demand that leads to a decline in sales revenue coupled with productivity investment that is not supported by profit maximization in equilibrium.

Table 2: Median Distribution of Revenue by Firm Size and Scale Adjustment in the Manufacturing Sector

	ln(Revenue)	**ln(Revenue)**		**ln(Revenue)**	
Year	**All Firms**	**Small Firms**	**Large Firms**	**Downsizers**	**Upsizers**
1994	11.01	7.15	12.21	**-**	11.01
1995	10.67	8.10	11.98	10.94	10.48
1996	9.92	9.06	11.99	9.99	9.89
1997	9.85	9.29	11.98	9.55	10.04
1998	10.11	9.49	12.88	9.25	10.67
1999	10.13	9.58	12.81	9.87	10.57
2000	10.34	9.73	12.45	10.00	10.87
2001	10.67	9.98	12.69	10.31	11.37
2002	10.85	10.05	12.66	10.62	11.56
2003	10.88	9.99	12.76	10.68	11.23
Median	10.42	**9.66**	**12.56**	**10.14**	**10.71**

Source: Author's Calculations

Table 2 reports the longitudinal log of sales revenue classified according to producers' scale of production and size adjustment; i.e., downsizers or upsizing firms referred to as firm-types. Notably, smaller plants and downsizers generated slightly lower revenues on a year-on-year basis compared to larger and upsizing firms. These performance patterns are consistent with findings in the international trade and industrial organization literature concerning the superiority of large and expanding firms (Haltiwanger 1997). In general, the revenue stream of all firms combined lingered around the median value overtime.

Table 3: Annual Distribution of Fully Flexible HARA Markups $(1+\mu_{it})$ and Technical Efficiency $(lnTFPQ)$ by Firm Size and Scale Adjustment

		Median Markups $(1+\mu_{it})$			
year	**Markups**	**Small Firms**	**Large Firms**	**Downsizers**	**Upsizers**
1994	1.75	1.75	2.52	.	1.75
1995	1.72	1.50	1.92	6.59	1.65
1996	1.54	1.47	1.65	3.26	1.49
1997	1.39	1.38	1.39	1.36	1.45
1998	1.49	1.48	1.93	1.55	1.49
1999	1.52	1.42	1.94	1.41	1.55
2000	1.49	1.43	1.85	1.55	1.47
2001	1.47	1.44	1.68	1.50	1.42
2002	1.49	1.43	1.70	1.44	1.56
2003	1.51	1.51	1.54	1.47	1.60
Median	**1.49**	**1.45**	**1.70**	**1.47**	**1.52**
		Median Technical Efficiency $(lnTFPQ_{it})$			
year	**Technical Efficiency**	**Small Firms**	**Large Firms**	**Downsizers**	**Upsizers**
1994	4.01	3.78	4.34	.	4.01
1995	3.96	3.77	4.15	5.69	3.91
1996	3.89	3.76	4.04	4.55	3.86
1997	3.77	3.74	3.98	3.75	3.85
1998	3.80	3.77	4.17	3.80	3.80
1999	3.79	3.75	4.14	3.75	3.84
2000	3.82	3.75	4.16	3.83	3.81
2001	3.82	3.79	3.95	3.83	3.81
2002	3.81	3.78	4.08	3.81	3.82
2003	3.78	3.77	3.99	3.78	3.79
Median	**3.81**	**3.76**	**4.06**	**3.80**	**3.82**

Source: Author's Calculations

Now one seeks to determine the scale elasticity in relation to the direction of markup flows and the behaviour of other market objects such as marginal costs (MC) vs average costs (AC) or scale economies vs diseconomies of scale and technical efficiency. Table 3 reports variable markups and technical efficiency. To illustrate the intuition of these numbers while preserving the simplicity of the exposition, here is a proof-of-concept example based on an identity in Syverson (2019) and Barkai (2020): $\mu = \frac{1}{1-S_H}$, where S_π is the profit share of revenue. Clearly, this heuristic identity shows that markups are an increasing function of profit-shares of revenue (S_π) as well as the scale elasticity (v). For instance, since the $\mu = 1.49$ and $S_\pi = 0.04$, then the scale elasticity is $v = (1 - 0.04) \times 1.49 = 1.44$.[14] The observed value of v simply means that the inverse elasticity of cost with respect to quantity up to the first-order moment (or the first derivative) of the cost technology is less than unity.[15] Apart from profit shares and scale elasticities, markups can also vary in response to variations in the output elasticity with respect to the freely variable input, i.e., material inputs. As a heuristic approach to high-level problem solving, the profit-share example enables us to make assumptions such as constant returns to scale (CRS) production technologies with a little more confidence.

However, Syverson (2019) warns that the practice of mixing statistical moments at different levels of granularity is not inconsequential since this ought to be a firm-to-firm relationship. Thus, given the data-driven scale economies from value-added production technologies $v =$ 0.94, one can infer that $MC \gtrsim AC$ since $\frac{AC}{MC} := v$ and $v \lesssim 1$. This means that an increase in MC relative to AC reflects a shift towards diseconomies of scale and deterioration in production efficiency, however measured. It is common wisdom that high marginal cost producers are also low productivity firms. As it turns out, the median $lnTFP^{rr}$ is such that $lnTFP^{rr}_{it} > lnTFP^{rr}_{it+1}$ $\forall t \in (1994, 2003)$ while $lnTFPQ_{it} > lnTFPQ_{it+1}$ $\forall t \in (1994, 1997)$. Notwithstanding pricing assumptions concerning the measurement of production efficiency, the marginal cost faced by firms persistently increased while markup prices fell.

Heuristics aside, a full structural demand configuration is reflected in the HARA parameter space $\{Q_{it}(\alpha, \mu, \rho) \colon \alpha \in (2.08,\ 9.75), \mu \in (0.15, 18.69)\ and\ \rho = 5.92\}$ which fixes ρ at its mid-point value of 5.92%. This demand structure produces a typically low-product complexity pattern of inverse demand elasticities, $\varepsilon^D_{FS0} \times \varepsilon^D_{FS1} = (-0.69,\ 0.01) \times (-0.54,\ 0.35)$ and also $\varepsilon^D_{FG0} \times \varepsilon^D_{FG1} = (-0.64,\ 0.01) \times (-0.69,\ 0.35)$, where FS0 vs. FS1 refer to SMEs vs. large firms and FG0 vs.FG1 refer to downsizers vs. upsizers, respectively. As expected, there are group-wise similarities in the demand for products. Both SMEs and downsizers faced similar consumer demand technologies with lower inverse elasticities relative to larger and expanding counterparts. For instance, producers of durable goods are either large or growing in size

[14] De Loecker et al. (2020) found that the average scale elasticity for their sample of US firms was $v = 1.03$ in 1980 and $v = 1.08$ in 2016, implying lower underlying marginal costs relative to average costs; and average costs were decreasing in quantity. Syverson (2004) reports $v = 0.996$ for the ready-made concrete, a homogeneous product variety with spatial and relational price variations, holding product differentiation in the quality dimension constant.

[15] In the special case where the production technology is homothetic, Syverson (2019) notes that v equates to returns to scale for that production function.

because they invest in productivity-enhancing activities. Consumer sensitivity to price adjustment for these products is naturally higher than for smaller and shrinking firms that produce largely primary commodities. One characteristic of a 'non-sophisticated' product is therefore a low inverse price elasticity of demand associated with low product prices.

As eyeball econometrics would have it, the first column shows a sharp decline in median markups during the first four years of trade liberalization and stabilizes thereafter. As it turns out, these patterns of longitudinal markup distribution are scale- and growth-dependent in that they also characterize firms regardless of size or whether firms are downsizing or upsizing their scale of operation. That is, firms had high but decreasing markups from the first several years of trade reforms. As expected, a similar trend obtained concerning technical efficiency. Thus, markups and productivity show some co-movement.

The use of first-order moments to summarize data is helpful for aggregate insights. However, to understand the patterns of heterogeneity around objects of interest, one can use firm-type classification to focus more on higher-order insights. Figure 4 presents visuals on cross-sectional distributions of markups and technical efficiency. The top left quadrant shows the distribution of markups for smaller firms (*FirmSize=0*) in the grey colour while the darker colour reflects the distribution of markups for larger firms (*FirmSize=1*). In general, the characterization of industrial markup distributions is also given by $skewness_{FS0} \times skewness_{FS1} = 1.43 \times 3.12$. Therefore, in spite of small-firm dominance over large firms in terms of observations, larger firms dominate the pricing power of firms as shown by the fat-right tail of the distribution. That is, large firms and a few downtown 'boutique' plants charged higher markups in the earlier years of trade reforms due to competitive effects of trade liberalization.[16] A similar pattern exists in the case of the distribution of markups by firm growth and contraction in the top right quadrant. Downsizing firms (*FirmGrowth=0*) dominated upsizing firms in number and also in markup pricing, especially in the earlier years, while upsizing firms with *FirmGrowth=1* retained a higher overall median markup. Again, although upsizing firms dominated downsizers in many dimensions, the $skewness_{FG0} \times skewness_{FG1} = 6.09 \times 5.01$.

[16] See Dhingra and Morrow (2019) for conditions surrounding small firms charging higher prices over marginal costs.

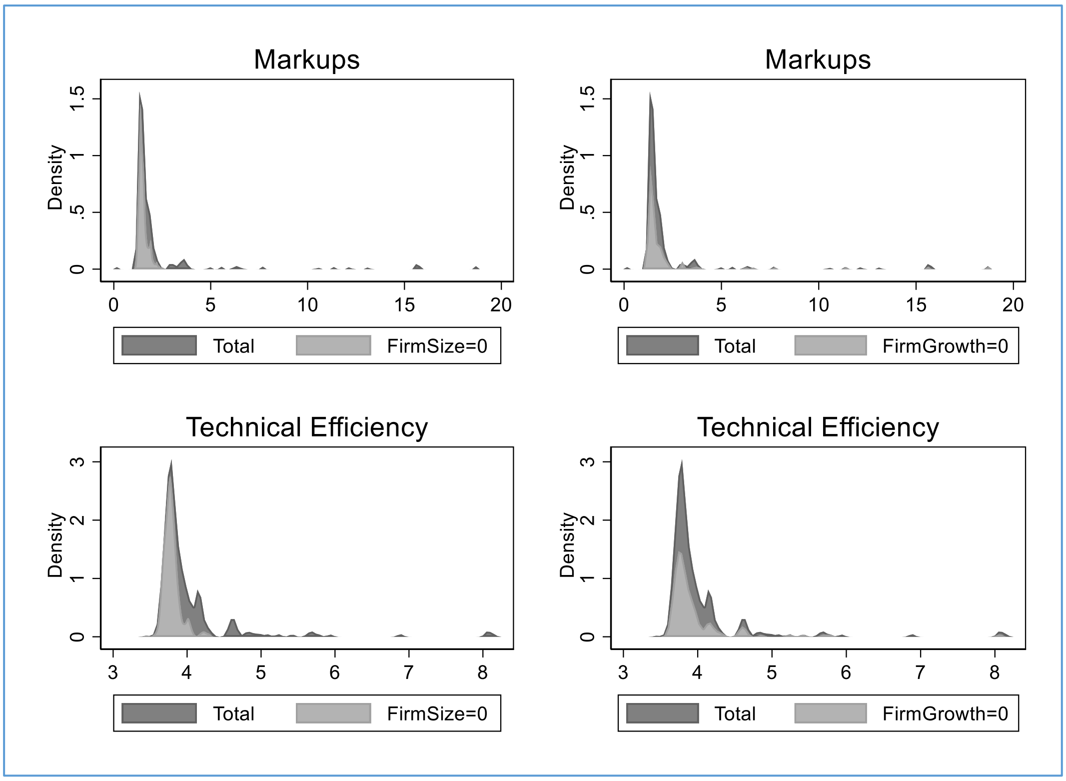

Figure 4: Markup and Productivity Distributions by Firm Size and Growth Patterns

Source: Author's Calculations.

Looking at the cross-section of productivity dynamics, similar second-order measures of production efficiency emerge. The size-dependent distribution of technical efficiency exhibits relatively more skewness for large plants (FS1) and downsizers (FS0) in the bottom panel: $skewness_{FS0} \times skewness_{FS1} = 1.42 \times 3.09$ and $skewness_{FG0} \times skewness_{FG1} = 5.23 \times 4.52$. Normally, firms shrink because of inefficiency in production and this seems not to be the case here. Firms appear to scale down operations in order to improve their production efficiency.

6.2. The Impact of Markups on Production Efficiency

The empirical analysis of the markup-productivity relationship relies on the presence of endogenous or exogenous structural effects of monopolistic markup pricing that drive variation in productivity. In the present context, markup endogeneity implies that there exists a set of high-dimensional instrumental variables, some constructed as series terms, which identify causal variation in markups but not in productivity. However, the use of the full set of IVs introduces model overfitting with adverse consequences on inference. To counteract model overfitting, we deployed model selection procedures using post-double selection methods based on Lasso to reduce model dimension through regularization as explained in Section 3. The same model architecture and deployment also applied under markup exogeneity assumptions.

Table 4 reports empirical results produced by relating markup pricing to three measures of production efficiency; namely, the baseline technical efficiency ($lnTFPQ^{HARA}$) and robustness checks of regression- based $lnTFP^{rr}$ and cost-share-based $lnTFP^{CS}$ measures of productivity. Under the assumption of previous period endogeneity of measured product-level markups, the First-Stage F-Statistic is 10.83, thereby validating the assumption that markups are endogenous. The related parameter estimate lies within the Wald-test confidence set with positive boundaries, and its magnitude suggests that a 10% increase in markups raises technical efficiency shocks by circa 5.43%.[17] That is, the standard incomplete pass-through of markups to productivity shocks is preserved. Although the FULLER test is by design robust to many instruments, it also generates an amplified *albeit* imprecisely measured coefficient. However, using the full set of controls significantly attenuates the order of magnitude for the coefficient on markups due to model overfitting and the wrong exogeneity assumption concerning covariates. Furthermore, the use of a full set of instrumental variables also introduces model overfitting. Evidently, the results derived from the full set of confounders and full set of IVs are uninformative for inference.

Table 4: The Impact of Markups on Productivity under HARA Pricing Assumptions

	Baseline Results	**Robustness Checks**	
Model Selection	$lnTFPQ^{HARA}$	$lnTFP^{rr}$	$lnTFP^{CS}$
Full Set of Controls	0.06 (0.04)	0.04*** (0.00)	0.82*** (0.07)
Exogenous Markups	0.34** (0.10)	0.14*** (0.01)	2.81*** (0.31)
Full Set of Instruments	0.08 (0.22)	0.04 (0.05)	-1.31 (0.52)
Endogenous Markups	0.54 *** (0.12)	-0.04 (0.06)	1.93 (2.62)
FULLER	0.63 (0.36)	0.097 (0.08)	1.63 (2.82)
Diagnostics for IVs			
Olea-Pflueger First-Stage F-Statistic	10.83	1.40	0.93

Notes: All regressions are revenue-weighted. Key: *p<0.05; **p<0.01; ***p<0.001.

Source: Author's Calculations.

If the market under investigation transitions from monopolistic competition with a steeper inverse demand curve that generates $lnTFPQ^{HARA}$ to a perfectly elastic inverse demand curve that generates $lnTFP^{rr}$ under perfect competition, we observe a breakdown in the weak IV rule of thumb associating endogeneity with the First-Stage F-Statistic that is greater than 10%. Given this outcome, it seems sensible to assume markup exogeneity in the perfectly competitive

[17] The markup-technical efficiency result mimics the findings of Cusolito (2017) for Chile, whose markup coefficient was 0.189, with confounders selected purely on the basis of economic intuition.

market. This choice of assumption produces significant and attenuated structural effects of markups on productivity. Furthermore, regression of the measured cost-share based production efficiency, $lnTFP^{CS}$ on exogenous markups generates a significant coefficient with an order of magnitude that is markedly amplified relative to the HARA measure of productivity. In both $lnTFP^{rr}$ and $lnTFP^{CS}$ robustness checks, the effects of markups on idiosyncratic productivity are broadly consistent with the baseline results.

However, the positive sign and statistical significance of the markup parameter may seem surprising to some readers and therefore requires some explaining. Typically, tougher competition introduced by trade liberalization in a heterogeneous-firm economy tends to reduce markups over marginal costs due to demand elasticity that falls with product sales (Mrázová and Neary, 2017). Therefore, an inverse markup-productivity relationship arises from the price increase relative to marginal costs that reduces productivity shocks (Aghion et al., 2008). In particular, higher aggregate productivity and lower average markups characterize larger and more integrated goods' markets *a priori* (Melitz and Ottaviano, 2008).

As a result, with average tariffs declining by 35.9% in the Customs Union, intermediate input tariffs fell more rapidly relative to those of outputs (Edwards and Behar, 2006; Edwards and Alves, 2006). The tariff adjustment had a direct downward effect on input prices. Perhaps because a lower Common External Tariff (CET) remained in force during trade reforms, export products originating within the Customs Union were largely intermediate inputs into the production of final goods in South Africa. With some investment in learning, quality upgrading, process innovation, and/or technology adoption in relation to intermediate inputs for market competitive; intermediate exports from Swaziland attracted higher $t-1$ markups relative to non-CU suppliers and translated into higher t productivity growth (cf. Edwards and Lawrence, 2006).

Turning to the comparative analysis of SMEs, large firms, downsizers and upsizing firms helps us understand the nature of structural effects of markups over hard-to-observe marginal costs on the technical efficiency in manufacturing. For each producer-type, we repeat the model selection procedure and apply Ordinary Least Squares (OLS) or Instrumental Variables (IVs), depending on empirical diagnostic results, to estimate the markup coefficient.

Table 5 reports the impact of markups on production efficiency, after controlling for firm-type. For each category of firms, we carry out estimation of the target parameter (γ) under exogeneity and endogeneity assumptions of markups. First and foremost, this estimation relies on the full set of covariates to reflect the extent of overfitting bias. Therefore, the value of the estimate under these conditions is not much of interest to analysts. Confronted with this bias, we implement regularization procedures for parsimonious model selection to facilitate unbiased estimation and inference. The choice between endogeneity and exogeneity of markups is influenced by diagnostic results of orthogonality conditions as in Baum et al. (2007). As discussed in the theoretical section, the validity of orthogonality conditions allows for the conduct of inference on the γ-coefficient based on the union of reduced form, first-stage and the amelioration set of covariates. However, when markup endogeneity assumptions hold; e.g., when the First-Stage Effective F-statistic by Montiel Olea-Pflueger (2013) exceeds the 10% threshold, then the standard Two-Stage-Least-Squares (TSLS) applies after model selection.[18] Therefore, all but for markups of small firms are endogenous.

Table 5: The Impact of Markups on Technical Efficiency by Firm-Type

Model Selection	**Small Firms**		**Large Firms**	
Full Set of Controls	-0.10***	(0.01)	0.04*	(0.02)
Exogenous Markups	0.05	(0.05)	0.32*	(0.11)
Full Set of Instruments	0.07	(0.05)	0.38	(0. 20)
Endogenous Markups	0.06	(0.05)	0.77***	(0.17)
FULLER	0.18	(0.17)	0.65*	(0.26)
Diagnostics for IVs				
Olea-Pflueger First-Stage F-Statistic	0.72		32.23	
	Downsizers ($\Delta E_{it} \leq 0$)		**Expanding Firms ($\Delta E_{it} > 0$)**	
Full Set of Controls	0.02	(0.07)	0.21***	(0.00)
Exogenous Markups	0.53***	(0.08)	0.39*	(0.13)
Full Set of Instruments	0.03	(0.05)	0.12	(0.06)
Endogenous Markups	0.54***	(0.04)	0.21***	(0.02)
FULLER	0.61***	(0.16)	-0.20	(0.57)
Diagnostics for IVs				
Olea-Pflueger First-Stage F-Statistic	19.85		54.27	

Notes: All regressions are revenue-weighted. Key: *p<0.05; **p<0.01; ***p<0.001.

Source: Author's Calculations.

The top-left quadrant reports the structural effects of markups on production efficiency. For firms employing less than 60 workers, the coefficient is insignificant at all sensible levels of confidence and this result is robust to markup endogeneity/exogeneity assumptions. Furthermore; even this *ad hoc* firm-size definition shows that $\frac{d\mu_{it}}{dt} < 0$ implies higher markups for small firms as depicted in Table 3 and figure 4, with the markup measure of skewness of 1.43. That is, a general decline in markups implies that small firms charge higher markups without

[18] Alternatively, one can also rely on weak-instrument robust inference to determine the assumption to adopt if the Effective F-statistic<10%.

statistically significant investment in productivity. This is consistent with the theoretical performance of 'downtown boutique' plants analysed by Dhingra and Morrow (2019) and Mhlanga (2023).

The rest of the firm-types have similar characteristics in the exogeneity/endogeneity conundrum. A unit percentage increase in the elasticity of markups under endogenous market structure increases the production efficiency by $0.00 \times 0.21 \times 0.54 \times 0.77$ for $SMEs \times Upsizers \times Downsizers \times Large\ Firms$. In Peters (2020) and Verhoogen (2023), such firms typically make investments in various dimensions of industrial upgrading to improve their technical efficiency in order to raise markups on existing product varieties. Alternatively, competitive effects of firm entry during trade liberalization may induce large multiproduct firms to skew their export product range towards a better performing product mix to increase technical efficiency (Mayer, Melitz and Ottaviano, 2014; 2021). However, there are two stylized facts associated with the latter case that require mentioning. First, the link between export demand shocks under imperfect competition and the product mix needs empirical validation. Second, the connection between the structural shocks of inverse demand and technical efficiency also calls for thorough investigation.

At least two reasons explain why firms downsize. First, plants scale down their workforce in the $\Delta E_{it} \leq 0$ firm-type because their productivity is too low and high productivity competitors drive the unproductive firms out of the market. The displaced workers are then absorbed by the productive establishments in the $\Delta E_{it} > 0$ type. Second, an alternative scenario is that producers downsize labour in order to be 'lean and mean' in technical efficiency. Again, labour displacement by downsizers avails new workers to upsizing firms. Perhaps the latter explanation is more plausible given the insignificant difference between the median productivities of $\Delta E_t \leq 0$ vs. $\Delta E_{it} > 0$ firm-types. Moreover, downsizers exhibited a circa 2.6 times impact of endogenous markups on idiosyncratic technical efficiency compared to upsizing firms. Thus, labour rationalization and potential investments in upgrading activities, broadly defined, strengthen the production efficiency of shrinking plants.

7. Summary and Conclusion

This article has explored the behavioural patterns of markups and productivity as well as the structural impact of endogenous markups on idiosyncratic production efficiency under conditions of monopolistic competition. Its main contribution lies in developing a high-dimensional sparse modelling framework that nests other regression-based methods for estimating the impact of endogenous market structure on granular production efficiency. By using B-splines, polynomials, their interactions and time trends, the procedure optimally selects exogenous covariates and instrumental variables for the causal identification of structural effects of markups on productivity.

The paper found pronounced markup variability among large firms and scale adjusting plants. It also found that the observed inelasticity of demand is a dominant characteristic for all firm-types; namely, SMEs, large firms, downsizers, and upsizing plants. Additionally, large and upsizing firms charge higher markups relative to downsizers and SMEs. However, downsizers are almost as productive as upsizing plants, indicating that the shrinking of establishments in this category was not just a corrective policy to poor performance but rather an investment in technical efficiency with the objective of raising markups in future.

The central finding of this study is that the impact of endogenous markups on idiosyncratic production efficiency is $\gamma \in [0.00, 0.77]$, conditional on firm-type. Thus, the structural effects of markups on production efficiency for large producers and downsizers are circa three times more than the markup coefficient for upsizing firms. The significant propensity to invest in production efficiency suggests that firms scaled down their workforce to enhance productivity growth rather than create a pathway to exit the market.

As is well known from Syverson (2011) and Verhoogen (2023), there are levers of productivity growth enhancement and also dimensions of industrial upgrading concerning existing products that may be necessary for public policy considerations. Equally importantly, product diversification into newer and more complex products may also be an area fruitful for public policy consideration, with special effort invested in attracting multiproduct firms that can participate in global value chains.

REFERENCES

Ackerberg D, Benkard L, Berry S, Pakes A. Econometric Tools for Analysing Market Outcomes. In: Heckman J, Leamer E: The Handbook of Econometrics. 6A. Amsterdam: North-Holland; 2007. pp. 4171-4276.

Ackerberg D, K. Caves, and G. Frazer (2015) Identification Properties of Recent Production Function Estimators, *Econometrica*, 83(6): 2411–2451. https://doi.org/10.3982/ECTA13408

Aghion, P., M. Braun, and J. Fedderke (2008) Competition and productivity growth in South Africa. *Economics of Transition* 16(4): 741-768. http://dx.doi.org/10.1111/j.1468-0351.2008.00336.x

Baker J. B. (2003) The Case for Antitrust Enforcement, *Journal of Economic Perspectives,* 17(4): 27–50. *http://DOI:10.1257/089533003772034880*

Barkai (2020) Declining Labour and Capital Shares; *The Journal of Finance*. https://DOI:10.1111/jofi.12909

Baum C. F., M. E. Schaffer and S. Stillman (2007) Enhanced Routines for Instrumental Variables/GMM Estimation and Testing, Boston College Economics Working Paper 667.

Beggs A. (2021) Demand functions and demand manifolds, *Economic Theory Bulletin*, 9:195–207. https://doi.org/10.1007/s40505-021-00203-1

Belloni A., D. Chen, V. Chernozhukov and C. Hansen (2012) Sparse Models and Methods for Optimal Instruments with An Application to Eminent Domain, Econometrica, 80 (6): 2369-2429. https://doi.org/10.3982/ECTA9626

Belloni A., V. Chernozhukov and C. Hansen (2014a) High-Dimensional Methods and Inference on Structural and Treatment Effects, *Journal of Economic Perspectives*, 28 (2): 29-50. *https://DOI:10.1257/jep.28.2.29*

Belloni A., V. Chernozhukov and C. Hansen (2014b) Inference on Treatment Effects after Selection among High-Dimensional Controls, *Review of Economic Studies*, 81: 608-650. https://doi.org/10.1093/restud/rdt044

Belloni A., V. Chernozhukov, C. Hansen, D. Kozbur (2016) Inference in High-Dimensional Panel Models with an Application to Gun Control, *Journal of Business & Economic Statistics*, 34 (4): 590-605. https://doi.org/10.1080/07350015.2015.1102733

Belloni A., V. Chernozhukov, I. Fernández-Val and C. Hansen (2017) Program Evaluation and Causal Inference with High-Dimensional Data, *Econometrica*, 85 (1): 233-298. https://doi.org/10.3982/ECTA12723

Belloni, A., and V. Chernozhukov (2013) Least Squares After Model Selection in High-Dimensional Sparse Models, *Bernoulli*, 19, 521-547. https://DOI:10.3150/11-BEJ410

Berry S.T. and G. Compiani (2021) Empirical Models of Industry Dynamics with Endogenous Market Structure, *Annual Review of Economics*, 13:309-334. https://doi.org/10.1146/annurev-economics-081720-120019

Bond S., A. Hashemi, G. Kaplan, and P. Zoch (2021) Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data, *Journal of Monetary Economics*, 121: 1–14. https://doi.org/10.1016/j.jmoneco.2021.05.004

Caner M. and A. B. Kock (2018) High Dimensional Linear GMM, arXiv:1811.08779v2.

Caner M. and H. H. Zhang (2014) Adaptive Elastic Net for Generalized Methods of Moments, *Journal of Business & Economic Statistics*, 32 (1): 30-47. https://doi.org/10.1080/07350015.2013.836104

Chen X. (2007), Large Sample Sieve Estimation of Semi -Nonparametric Models, Handbook of Econometrics, 6: 5559-5632.

Chen X. and D. Pouzo (2012) Estimation of Nonparametric Conditional Moment Models with Possibly Nonsmooth Moments, *Econometrica* , 80: 277-322. https://doi.org/10.3982/ECTA7888

Cusolito A.P., A. Garcia-Marin and W. F. Maloney (2017), Competition, Innovation and Within-Plant Productivity: Evidence from Chilean Plants, Mimeograph.
De Loecker J., J. Eeckhout, and G. Unger (2020). The Rise of Market Power and The Macroeconomic Implications. *The Quarterly Journal of Economics*, 135 (2), 561–644. https://doi.org/10.1093/qje/qjz041.
De Loecker J., P. K. Goldberg, A. K. Khandelwal, and N. Pavcnik (2016). Prices, Markups, and Trade Reform. *Econometrica*, 84 (2), 445–510. https://doi.org/10.3982/ECTA11042
De Loecker, J. and F. Warzynski (2012) Markups and Firm-Level Export Status, *American Economic Review*, 102 (6): 2437–2471. *https://DOI:10.1257/aer.102.6.2437*
Demirer M. (2022). Production Function Estimation with Factor-Augmenting Technology: An Application to Markups. MIT Sloan School of Management.
Dhingra S. and J. Morrow (2019). Monopolistic Competition and Optimum Product Diversity under Firm Heterogeneity. *Journal of Political Economy,* 127 (1): 196-232. https://doi.org/10.1086/700732
Edwards L. and A. Behar (2006) Trade Liberalisation and Labour Demand within South African Manufacturing Firms, *Studies in Economics and Econometrics*, 30 (2): 127-146. https://doi.org/10.1080/10800379.2006.12106411
Edwards L. and P. Alves (2006) South Africa's Export Performance: Determinants of Export Supply, *South African Journal of Economics*, 74 (3): 473-500. https://doi.org/10.1111/j.1813-6982.2006.00087.x
Foster L. S., C. A. Grim, J. Haltiwanger and Z. Wolf (2017). Macro and Micro Dynamics of Productivity: From Devilish Details to Insights, NBER Working Paper Series.
Foster, L., Grim, C., Haltiwanger, J. and Wolf, Z. (2016). Firm-Level Dispersion in Productivity: Is the Devil in the Details? *The American Economic Review*, 106(5): 95-98. https://doi.org/10.1257/aer.p20161023
Foster, L., R. Kulick and C. Syverson (2018). Misallocation Measures: The Distortion That Ate the Residual. NBER Working Paper 24199.
Haltiwanger J (1997) Measuring and Analysing Aggregate Fluctuations: The Importance of Building from Microeconomic Evidence, *Review*: 55-77.
Haltiwanger J. C. and R. W. Cooper (2006) On the Nature of Capital Adjustment Costs, *The Review of Economic Studies*, 73 (3): 611-633. https://doi.org/10.1111/j.1467-937X.2006.00389.x
Haltiwanger J., Kulick R., and Syverson C. (2018). Misallocation Measures: The Distortion That Ate the Residual, NBER, Cambridge.
Hausmann R., J. Hwang, and D, Rodrik (2007) What you export matters, Journal of Economic Growth, 12:1–25. https://doi:10.1007/s10887-006-9009-4
Hidalgo C. A., B. Klinger, A.-L. Barabási and R. Hausmann (2007) The Product Space Conditions the Development of Nations, *Science*, 317 (5837): 482-487.
https://doi.org/10.1007/s40505-021-00203-1
Hu Y., G. Huang and Y. Sasaki (2023) robustpf: A command for robust estimation of production functions, *The Stata Journal*, **23** **(**1): 86–96. https://doi:10.1177/1536867X231161977
Javanmard A. and A. Montanari (2014), Confidence Intervals and Hypothesis Testing for High-Dimensional Regression, *Journal of Machine Learning Research*, 15: 2869-2909.
José Luis Montiel Olea and C. Pflueger (2013) A Robust Test for Weak Instruments, *Journal of Business & Economic Statistics,* 31 (3): 358-369. https://doi.org/10.1080/00401706.2013.806694
Kock A. B. (2016) Oracle Inequalities for High-Dimensional Panel Data, Aarhus University and Creates, mimeograph.
Kock A. B. and H. Tang (2019) Uniform Inference in High-Dimensional Dynamic Panel Data Models with Approximately Sparse Fixed Effects, *Econometric Theory*, 35: 295–359. https://doi:10.1017/S0266466618000087

Lazear E. P. (1990) Job Security Provisions and Employment, *The Quarterly Journal of Economics*, 105 (3): 699-726. https://doi.org/10.2307/2937895

Levinsohn, J. and A. Petrin (2003) Estimating Production Functions Using Inputs to Control for Unobservables, *Review of Economic Studies,* 7: 317-341. https://doi.org/10.1111/1467-937X.00246

Mayer T., M. J. Melitz, and G. I. P. Ottaviano (2014) Market Size, Competition, and the Product Mix of Exporters, *American Economic Review*, 104(2): 495-536. http://dx.doi.org/10.1257/aer.104.2A95

Mayer T., M. J. Melitz, and G. I. P. Ottaviano (2021) Product Mix and Firm Productivity Responses to Trade Competition, *The Review of Economics and Statistics*, 103(5): 874–891. https://doi.org/10.1162/rest_a_00952

Melitz M. J. and G. I. P. Ottaviano (2008) Market Size, Trade, and Productivity, *Review of Economic Studies,* 75: 295–316. https://doi.org/10.1111/j.1467-937X.2007.00463.x

Mhlanga S. V. (2023) Productivity and Reallocation under Monopolistic Competition, with Distortions: Micro Panel Data Analysis, *South African Journal of Economics.* 91(4): 466-497. https://doi.org/10.1111/saje.12355

Mhlanga S. V. and N. R. Rankin (2021). Fixed Costs, Markups and Concentration in Eswatini Swaziland): A Firm-Level Analysis of Panel Data. *South African Journal of Economics*, 89, 391-416. https://doi.org/10.1111/saje.12289.

Mrázová M. and J. P. Neary (2017). Not So Demanding: Demand Structure and Firm Behaviour. *American Economic Review,* 107(12): 3835–3874. http://dx.doi.10.1257/aer.20160175.

Mrázová M., J. P. Neary, and M. Parenti (2021). Sales and Markup Dispersion: Theory and Empirics. *Econometrica*, 89 (4), 1753-1788. https://doi.org/10.3982/ECTA17416.

Newey W. K. (1997) Convergence Rates and Asymptotic Normality for Series Estimators, *Journal of Econometrics*, 79: 147-168. https://doi.org/10.1016/S0304-4076(97)00011-0

Olley, S., and A. Pakes (1996). The Dynamics of Productivity in The Telecommunications Equipment Industry. *Econometrica,* 64, 1263– 1298. https://doi.org/10.2307/2171831.

on Structural and Treatment Effects, *Journal of Economic Perspectives, 28 (2): 29–50. DOI: https://doi.org/10.1257/jep.28.2.29*

Perets G. S. and E. Yashiv (2015). The Fundamental Nature of HARA Utility. Discussion Papers 1522, Centre for Macroeconomics (CFM).

Peters M. (2020). Heterogeneous Markups, Growth, and Endogenous Misallocation. *Econometrica*, 88 (5), 2037–2073. https://doi.org/10.3982/ECTA15565.

Rovigatti G. and V. Mollisi (2018) Theory and practice of total-factor productivity estimation: The control function approach using Stata, *The Stata Journal*, Vol. 18 (3): 618-662. https://doi.org/10.1177/1536867X1801800307

Syverson C. (2004) Product Substitutability and Productivity Dispersion, *The Review of Economics and Statistics*, 86 (2): 534-550. http://dx.doi.org/10.1162/003465304323031094

Syverson C. (2011) What Determines Productivity? *Journal of Economic Literature, 49:2, 326–365. https://doi.org/10.1257/jel.49.2.326*

Syverson C. (2019) Macroeconomics and Market Power: Context, Implications, and Open Questions, *Journal of Economic Perspectives,* 33 (3): 23–43. https://doi.org/10.1257/jep.33.3.23

Verhoogen E. (2023) Firm-Level Upgrading in Developing Countries, *Journal of Economic Literature*, 61(4): 1410–1464. https://doi.org/10.1257/jel.20221633

Wooldridge, J.M. (2009) On estimating firm-level production functions using proxy variables to control for unobservables, *Economics Letters*, 104: 112–114. https://doi.org/10.1016/j.econlet.2009.04.026

Zhang C-H. and S. S. Zhang (2014), Confidence intervals for low dimensional parameters in high dimensional linear models, *Journal of the Royal Statistical Society.* 76 (1): 217–242. https://doi.org/10.1111/rssb.12026