Production functions, markups, and the limits of the World Bank enterprise surveys: what can be identified from Low-Income firm surveys?

Alex Okolouma — University of Dschang — okoloumalex@yahoo.fr

September 2026

Abstract

Firm-level markups and total factor productivity are now routinely estimated for developing economies from repeated establishment surveys, above all the World Bank Enterprise Surveys (WBES). This paper asks what those data can actually identify. Using 20,047 firms across 24 Sub-Saharan African economies, I show that the standard cross-sectional route delivers an output elasticity of materials of 0.64 and a median De Loecker–Warzynski markup of 2.1 — a level no plausible model of African manufacturing supports. I then eliminate the obvious culprit: broadening the variable-input aggregate from raw materials to materials plus energy, water and other production costs moves the median markup only from 2.11 to 1.99, because the estimated elasticity rises almost proportionally with the aggregate. The problem lies in the elasticity, not the input measure. Assembling every WBES panel release for the region (982 firms in 14 economies observed twice) I show that proxy-variable estimators (Olley–Pakes, Levinsohn–Petrin, Ackerberg–Caves–Frazer) are unavailable in principle, not merely underpowered: waves are separated by 3 to 18 years, so the first-order Markov assumption on productivity fails by construction. First-differencing lowers the elasticity to 0.54 and the implied markup to 1.68, close to an independent accounting benchmark of 1.81. That agreement is not identification. Exploiting Kenya's four-wave panel, whose firms span many differencing intervals *within one country*, I show the first-difference estimate is itself attenuated: the elasticity rises from 0.13 at a five-year interval to 0.68 at seven-to-twelve years, with an interaction significant at $p = 0.013$. Ordinary least squares and first differences therefore bracket the parameter rather than either recovering it. The paper's contribution is a design (within-country variation in the differencing interval) that turns the irregular wave spacing of firm surveys from a nuisance into a diagnostic, and a partial identification result: the materials elasticity lies in [0.54, 0.75] and the median markup in [1.68, 2.31], with the sign and approximate magnitude of the bias at each bound known. A correlation analysis further shows that estimated markups and estimated productivity are two transformations of variation in a single reported ratio, so measurement error in one survey question propagates, with opposite sign, into both headline quantities. These limits bear directly on the industrialisation diagnosis: manufacturing fell from 13.9 to 9.9 per cent of Sub-Saharan African GDP between 2000 and 2024, and choosing among the firm-level explanations for that decline requires levels these data do not deliver.

Keywords: production function estimation, markups, measurement error, partial identification, firm surveys, Sub-Saharan Africa

JEL: C81, D24, L11, O14, O55

1. Introduction

1.1 Why the measurement question matters

Sub-Saharan Africa is not industrialising, and the aggregate record is unambiguous. Manufacturing value added fell from **13.95 per cent of regional GDP in 2000 to 9.93 per cent in 2024**; the world figure fell over the same period from 18.41 to 14.89 per cent, so the region did not merely track a global trend but lost ground from a lower base (World Bank, 2025). This is the pattern Rodrik (2016) labelled premature deindustrialisation: economies shedding manufacturing shares at income levels far below those at which today's advanced economies did so.

The policy architecture points the other way. Sustainable Development Goal target 9.2 commits signatories to raise industry's share of employment and GDP significantly and to **double it in least developed countries** by 2030, monitored through indicator 9.2.1, manufacturing value added as a proportion of GDP (United Nations, 2015). The African Union's Agenda 2063 is more ambitious still: Goal 4, *Transformed Economies*, targets a manufacturing sector several times its 2013 real value. Assessments of the first implementation decade rank Goal 4 the **worst performing of the twenty goals**, at roughly 17 per cent implementation (Mo Ibrahim Foundation, 2024). The gap between target and outcome is therefore large, widening, and politically salient.

Diagnosing that gap requires knowing what is happening inside firms. An aggregate manufacturing share is an accounting identity over establishments; whether it falls because firms are unproductive, because inputs are misallocated across them, or because market power lets incumbents restrict output are different diagnoses with different remedies. Each of those diagnoses is empirically expressed as a firm-level productivity or markup estimate, and in this region such estimates come, almost without exception, from one source: the World Bank Enterprise Surveys.

That is the dependency this paper examines. If firm-level markups and total factor productivity cannot be identified from these surveys, then the microeconomic diagnosis underpinning industrial policy in the region rests on quantities the data do not deliver — and the failure is invisible, because the estimators return plausible-looking numbers either way.

1.2 The argument

The estimation of firm-level production functions has a canonical solution. Because unobserved productivity is known to the firm when it chooses inputs, ordinary least squares confounds technology with input choice — the transmission bias identified by Marschak and Andrews (1944). Olley and Pakes (1996), Levinsohn and Petrin (2003) and Ackerberg, Caves and Frazer (2015) resolve it by using an observed decision — investment or intermediate input demand — as a proxy that inverts to recover productivity, and by imposing that productivity follows a first-order Markov process so that lagged inputs are valid instruments for the innovation. De Loecker and Warzynski (2012) then convert the estimated output elasticity of a flexible input into a firm-level markup, dividing it by that input's share in revenue. This machinery underpins a large

empirical literature, including the claim that markups have risen substantially in recent decades (De Loecker, Eeckhout and Unger, 2020).

Applying it to low-income economies is attractive and increasingly common, because the World Bank Enterprise Surveys provide harmonised firm microdata for more than 150 countries, including nearly every economy in Sub-Saharan Africa, with sales, employment, labour cost, material inputs and capital expenditure recorded on a common instrument. The World Bank itself estimates Cobb–Douglas and translog production functions from these data and publishes firm-level total factor productivity.

This paper asks a prior question: *what can these data identify?* The answer developed here is narrower than the literature's practice assumes, and the way it is narrow is informative.

I proceed by elimination, and I report the eliminations rather than only the survivor, because two of the three natural diagnoses turn out to be wrong and their failure is what locates the real problem.

The cross-sectional benchmark, estimated on 5,789 manufacturing firms pooled across 22 economies with country-by-year and revision-consistent industry fixed effects, yields plausible-looking elasticities — 0.46 on labour, 0.64 on materials — but a median markup of 2.11. A markup of two implies price at twice marginal cost across the median African manufacturer, which no serious model of these markets supports, and the distribution's mean of roughly 18 is plainly an artefact.

The obvious explanation is that the survey's single materials question understates intermediate inputs, deflating the denominator of the markup formula. Section 4 tests this and **rejects it**. Adding electricity and fuel, then water, internet and other production costs, raises the median input share from 0.30 to 0.36 — but the estimated elasticity rises in step, from 0.643 to 0.726, so the median markup falls only from 2.11 to 1.99. The ratio is close to invariant to the input aggregate, because the regression re-estimates the numerator on whatever denominator is chosen. The defect is in the elasticity.

Section 5 introduces an accounting benchmark that requires no production function at all: revenue over total production cost, available directly in six economies, has a median of 1.36, and revenue over variable cost a median of 1.66 to 1.81 depending on sample. These are mutually consistent — the markup over variable cost must exceed the markup over total cost — and they bracket a plausible region. The structural estimate sits far above both.

If the cross-sectional elasticity is biased upward by transmission, the textbook remedy is a proxy-variable estimator on panel data. Section 6 assembles every WBES panel release for Sub-Saharan Africa: 22 releases, 16 economies, and after imposing that the production inputs be complete in both waves, 982 firms across 14 economies. The estimator is nonetheless unavailable, for a structural rather than statistical reason. WBES panels observe firms twice, and the two waves are separated by between three and eighteen years, with a median of six. The proxy-variable framework requires productivity to follow a first-order Markov process *at the observation frequency*; over a six-year interval the innovation term absorbs essentially all of

productivity, and the lagged input loses its instrumental content. No increase in sample size repairs this.

What two observations per firm do permit is first-differencing, which sweeps out the permanent component of firm productivity — the dominant source of transmission bias. Section 7 reports the result: the materials elasticity falls from 0.748 in pooled OLS on the same firms to 0.542, robust to a wild cluster bootstrap over 14 economies ($p = 0.0004$). The implied markup falls to 1.68, within a few percent of the accounting benchmark of 1.81.

Section 8 shows this agreement is a coincidence of opposing errors. Differencing removes transmission bias but amplifies attenuation from measurement error, a trade-off analysed by Griliches and Hausman (1986), who observed that the degree of attenuation depends on the length of the difference taken. Their insight is directly testable here, because WBES waves are irregularly spaced. The cross-country version of the test is confounded — the differencing interval is a country attribute, so a gap split is a country split — and returns an insignificant interaction. Kenya's four-wave panel breaks the confound: its 224 firms span intervals from five to eighteen years *within a single economy*. There, the elasticity rises monotonically from 0.133 at five years, where it is statistically indistinguishable from zero and implies an impossible markup of 0.41, to 0.683 at seven to twelve years, with the interaction significant at $p = 0.013$.

Attenuation is therefore real and quantitatively serious, and the first-difference estimate is a lower bound rather than a correction. Since OLS is biased upward by transmission and first differences downward by measurement error, the two **bracket** the parameter. Section 9 states the resulting partial identification result: the materials elasticity lies in [0.54, 0.75] and the median markup in [1.68, 2.31]. This is far more informative than the Klepper–Leamer (1984) reverse-regression bound, which on these data admits elasticities up to 1.71 and is therefore vacuous.

The contribution is threefold. Substantively, markup and productivity *levels* are not identified from WBES, and published estimates that report them should be read as bounds at best. Methodologically, the within-country differencing-interval design converts the irregular wave spacing of firm surveys — normally a nuisance — into a diagnostic for measurement error, and it transfers to any repeated firm survey with uneven waves. Practically, the paper documents what the data do and do not contain, including several traps that silently corrupt analyses built on them.

2. Data

2.1 Source and construction

The World Bank Enterprise Surveys are stratified random samples of the registered, non-agricultural private economy, with strata on subnational region, sector and size. I harmonise every available Sub-Saharan African wave into a single file, resolving the variable-name drift that occurs across survey generations. The pooled dataset contains **20,047 firms across 24 economies and 47 economy-years**.

Production function estimation requires output, labour, intermediate inputs and capital. WBES supplies total annual sales, permanent full-time employment with a production/non-production split, total labour cost, the cost of raw materials and intermediate goods, electricity and fuel costs, and expenditure on equipment and on land and buildings. It also records sales and employment three years prior, giving a coarse retrospective dimension within a single cross-section.

Two features of this list matter for what follows.

First, the materials question is administered only in the manufacturing module. The subsample with sales, materials and labour cost jointly present is therefore **5,870 firms across 40 economy-years and 22 economies** — effectively the manufacturing sample, which is the appropriate estimation universe.

Second, **capital is measured as expenditure, not as a stock**. WBES records annual purchases of equipment and of land and buildings, not a perpetual-inventory capital stock. Any capital elasticity estimated from these data measures the contribution of current investment, which is a different object. I report specifications with and without capital and do not rest any conclusion on the capital coefficient.

2.2 Industry classification

Sector is recorded in a four-digit ISIC code whose revision changed across survey generations: earlier waves use one revision, waves from the 2022 methodology another. The same two-digit code denotes different industries in the two revisions. Pooling them silently — which the variable names invite, since the field is simply the sector code — mixes unrelated industries inside a single fixed effect. I therefore tag each observation with its revision and construct industry fixed effects on the revision–code pair, so codes from different revisions are never pooled.

2.3 Variables and their provenance

Table 1 sets out every variable used, its survey code, and the study that established its role in this literature. The mapping is not mechanical: which input is treated as flexible, and therefore which one carries the markup, is a modelling decision inherited from De Loecker and Warzynski (2012), while the choice of materials as the proxy that inverts to recover productivity comes from Levinsohn and Petrin (2003) and its correction in Ackerberg, Caves and Frazer (2015).

Table 1 — Variables, survey codes, and their role in the literature

Variable	WBES code	Definition	Role, after
Output y	d2	Total annual sales, last fiscal year	Gross output; revenue-based caveat: Klette and Griliches (1996)
Labour l	l1, l3a/l3b	Permanent full-time	Free input: Olley and

Variable	WBES code	Definition	Role, after
		employees; production / non-production split	Pakes (1996)
Materials m	n2e	Cost of raw materials and intermediate goods used in production	Flexible input carrying the markup: De Loecker and Warzynski (2012); proxy: Levinsohn and Petrin (2003), Ackerberg, Caves and Frazer (2015)
Labour cost	n2a	Total wage bill including bonuses	Cost-share benchmark: Foster, Haltiwanger and Syverson (2008)
Capital k	n5a + n5b	Expenditure on equipment; on land and buildings	Investment proxy: Olley and Pakes (1996). **Not a stock**
Energy	n2b, n2f	Annual cost of electricity; of fuel	Input-aggregate robustness (Section 4)
Total production cost	n2p	Total cost of production (6 economies)	Accounting markup benchmark (Section 5)
Retrospective output	n3, l2	Sales and employment three years prior	Within-firm dynamics where no panel exists
Industry	d1a2 / d1a2_v4	Four-digit ISIC, revision-dependent	Fixed effects (Section 2.2)
Sampling weight	wstrict	Strict stratification weight	Population-representative descriptives

Notes. Survey codes follow the harmonised World Bank Enterprise Surveys Global Methodology. Codes marked with a revision suffix changed across survey generations; see Section 2.2. The markup is a derived quantity, not an estimated coefficient.

The dependent variable throughout is log sales; the independent variables of interest are log employment and log materials, with log capital expenditure entering secondary specifications. The markup is a derived quantity, not an estimated coefficient: it is the ratio of the estimated materials elasticity to the observed materials share of revenue.

2.4 Correlation structure

Table 2 reports pairwise correlations on the manufacturing estimation sample. Three features bear on the results that follow.

Table 2 — Correlation matrix, manufacturing estimation sample (N = 5,789; capital N = 2,338)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) log sales	1.00							
(2) log employment	0.50	1.00						
(3) log materials	0.88	0.47	1.00					
(4) log wage bill	0.90	0.49	0.89	1.00				
(5) log capital expenditure	0.80	0.42	0.76	0.80	1.00			
(6) log labour productivity	0.90	0.07	0.78	0.78	0.72	1.00		
(7) materials share of revenue	0.02	0.03	0.39	0.11	0.04	0.01	1.00	
(8) log TFP (Cobb–Douglas residual)	0.64	−0.04	0.24	0.42	0.44	0.76	−0.50	1.00

Notes. Pairwise Pearson correlations on the manufacturing estimation sample. N = 5,789 for all pairs except those involving capital expenditure, where N = 2,338 (only firms reporting positive capital expenditure). Log TFP is the Cobb–Douglas residual from Table 3, column (1); its −0.50 correlation with the materials share is mechanical and is discussed in the text.

First, **log materials and log sales correlate at 0.88**, and materials with the wage bill at 0.89. The right-hand side is highly collinear, which is ordinary in production data but means the split of explanatory power between labour and materials is fragile — a fragility that surfaces directly in Section 7, where pooled OLS assigns labour an implausibly small coefficient.

Second, **the materials share is essentially uncorrelated with firm size** (0.02 with log sales, 0.03 with log employment). Whatever drives variation in the share, it is not a size or scale artefact, which rules out the simplest composition explanation for the implausible markups of Section 3.

Third, and most consequential: **the materials share correlates −0.50 with estimated log TFP**. This is mechanical rather than economic. The Cobb–Douglas residual subtracts $\beta_M m$ from output, so a firm reporting a low materials share is assigned high residual productivity; the same low share, in the denominator of the markup formula, is assigned a high markup. **Estimated markups and estimated TFP are therefore not two independent readings of firm performance — they are two transformations of the same residual variation in one reported ratio.** Any study using these data to relate markups to productivity risks recovering this identity rather than a behavioural relationship. The point is not incidental to what follows: it means measurement error in a single survey question propagates, with opposite sign, into both headline quantities.

3. The cross-sectional benchmark

3.1 Specification

I estimate a gross-output production function in logarithms,

$$y_i = \beta_L \, l_i + \beta_M \, m_i + \beta_K \, k_i + \delta_{ct} + \delta_s + \varepsilon_i,$$

where y is log sales, l log employment, m log materials and k log capital expenditure; δ_{ct} are country-by-year fixed effects, absorbing currency units, price levels and macroeconomic conditions, and δ_s are revision-consistent two-digit industry effects. Standard errors are clustered by country-year. Because country-by-year effects absorb price levels additively in logs, no separate deflation is applied; the estimates are revenue-based throughout, a limitation discussed in Section 9.

The De Loecker–Warzynski markup follows from cost minimisation in a flexible input:

$$\mu_i \; = \; \theta_M / \frac{P_i^M M_i}{P_i Q_i},$$

the output elasticity of materials divided by the materials share of revenue.

3.2 Results

On 5,789 manufacturing firms the Cobb–Douglas specification returns $\beta_L = 0.458$ (s.e. 0.047) and $\beta_M = 0.643$ (0.026). Adding the capital expenditure proxy, on the 2,333 firms reporting positive capital expenditure, gives $\beta_L = 0.424$, $\beta_M = 0.586$ and $\beta_K = 0.117$ (0.027). A translog

specification, which permits firm-specific elasticities, produces an average materials elasticity close to the Cobb–Douglas value.

Taken as technology parameters these are unremarkable. The markups they imply are not.

Table 3 — Gross-output production function and implied markups, cross-section

	(1) Cobb–Douglas	(2) + capital expenditure
log employment	0.458***	0.424***
	(0.047)	(0.048)
log materials	0.643***	0.586***
	(0.026)	(0.038)
log capital expenditure	—	0.117***
		(0.027)
Implied median markup	2.11	1.83
Country × year FE	Yes	Yes
Industry FE (revision-tagged)	Yes	Yes
Observations	5,782	2,333
Economy-year clusters	40	38

Notes. Dependent variable is log annual sales. Manufacturing firms with positive sales, employment, materials and labour cost. Standard errors, in parentheses, are clustered by economy-year. Column (2) is restricted to firms reporting positive capital expenditure. The implied markup is the estimated materials elasticity divided by the sample median materials share of revenue (0.30). *** p<0.01, ** p<0.05, * p<0.1.

The median firm is estimated to price at more than twice marginal cost, the interquartile range extends beyond four, and the mean exceeds 16. Dispersion in log total factor productivity within economy-year cells has a median standard deviation of 0.99, which is large even by the standards of the misallocation literature (Hsieh and Klenow, 2009).

The cross-economy ordering is more defensible than the levels. Median markups are highest in Angola, the Central African Republic, South Sudan and the Democratic Republic of Congo — small, fragmented, conflict- or resource-affected markets — and lowest in Senegal, South Africa, Benin and Burundi. The ranking is economically sensible. The levels are not credible, and it is the levels that the method purports to deliver.

Figure 1 makes the problem visible. The structural markup distribution lies to the right of the accounting benchmark of Section 5 across essentially its whole range, and **6.4 per cent of firms are assigned a markup below one** — pricing below marginal cost, which cost minimisation excludes by construction. A distribution placing one firm in fifteen in an impossible region is not mismeasuring a few outliers; it is misplaced as a whole.

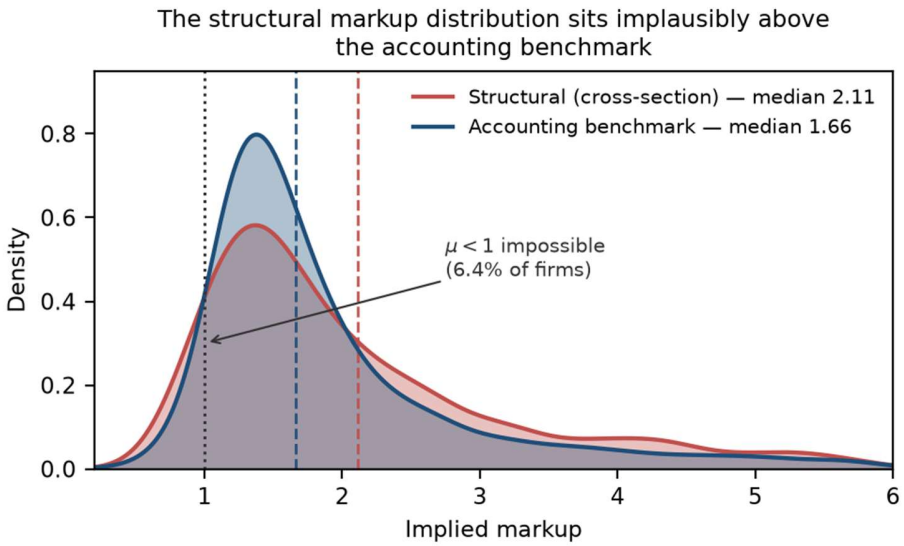


Figure 1 — Implied markup distributions. *Gaussian kernel densities on the manufacturing estimation sample, top percentile trimmed. Dashed vertical lines mark medians; the dotted line at unity marks the boundary below which a markup is economically impossible.*

4. Is the problem the input measure? A test, and a rejection

The immediate suspicion is arithmetic. The markup is an elasticity divided by an input's revenue share, and the measured share is low: the median ratio of materials to sales is **0.30**, against 0.5 to 0.65 in manufacturing censuses. The WBES materials question plausibly omits purchased services and other intermediates. A denominator too small by a third would inflate markups by half.

This hypothesis is testable, because WBES itemises other production costs. I construct three nested variable-input aggregates and re-estimate the production function on each, so that numerator and denominator always refer to the same input. Table 4 reports the result.

Table 4 — Does a broader variable-input aggregate rescue the markup?

Variable-input aggregate	Median revenue share	Elasticity θ_V	Implied median markup	N
Raw materials only	0.30	0.643***	2.11	5,782
		(0.026)		
+ electricity and fuel	0.35	0.721***	2.04	5,768
		(0.025)		
+ water, internet, other	0.36	0.726***	1.99	5,756

Variable-input aggregate	Median revenue share	Elasticity θ_V	Implied median markup	N
		(0.024)		

Notes. Each row re-estimates the production function on the input aggregate named, so numerator and denominator of the markup always refer to the same input. Specification, sample and clustering as in Table 3, column (1). *** p<0.01, ** p<0.05, * p<0.1.

The hypothesis fails. Broadening the aggregate raises the share by a fifth, but the estimated elasticity rises almost proportionally, and the median markup falls only from 2.11 to 1.99 — nowhere near the plausible range. The ratio is close to invariant to the choice of aggregate, for a mechanical reason: the regression re-estimates the elasticity on whichever input is used, so numerator and denominator move together.

This is worth stating positively. **Input mis-measurement is not the binding problem.** The defect is in the estimated elasticity itself, which sends the diagnosis toward estimation rather than instrumentation of the survey.

5. An accounting benchmark

Before repairing the elasticity it helps to know what a plausible markup looks like in these data, using no production function at all. Under cost minimisation the ratio of revenue to variable cost is a markup over variable cost, and where a total production cost is reported directly, revenue over that cost is a markup over total cost. WBES records a total production cost in six economies. Table 5 reports both measures.

Table 5 — Accounting markups (no production function estimated)

Measure	Economies	N	p25	Median	p75
Revenue / variable cost (intermediates + labour)	22	5,731	1.32	**1.66**	2.65
Revenue / total production cost (n2p)	6	609	1.14	**1.36**	1.90

Notes. Ratios computed firm by firm, top percentile trimmed. Both impose cost minimisation and constant returns and ignore the user cost of capital; they are a reference region, not an estimate of the markup. The second row is available only where the survey records a total production cost.

The two are mutually coherent: a markup over variable cost must exceed one over total cost, and 1.66 > 1.36. Both sit in a plausible region. The structural estimate of 2.1 sits well above both, and the gap is the object to be explained.

These benchmarks are not themselves estimates of the markup — they impose constant returns and ignore the user cost of capital. They are a reference region, and that is how they are used below.

6. The panel, and why proxy-variable estimators are unavailable

6.1 Assembly

The World Bank publishes dedicated panel releases for economies where firms were re-interviewed. I retrieved every such release for Sub-Saharan Africa — 22 releases, 18 usable files, 16 economies — and linked firms within each.

Two properties of these files defeat naive handling and are worth recording, because both silently destroy sample rather than raising an error. First, the linking key is sometimes named for the earlier wave (`idpanel2007`) rather than generically (`panelid`); matching only the generic name makes six economies, including Kenya, Senegal, Tanzania and Uganda, appear to be in an unusable format. Second, at least one file (Uganda) carries **both** a generic key that links nothing and a wave-named key that links correctly; selecting the key by name order silently drops the economy. The key must be chosen by which candidate actually links waves. Together these cost 245 firms and four economies before being caught.

Requiring sales, employment and materials to be present in both waves yields the estimation sample: **982 firms, 2,001 firm-years, 14 economies**.

6.2 Why Olley–Pakes, Levinsohn–Petrin and Ackerberg–Caves–Frazer do not apply

The proxy-variable framework rests on productivity following a first-order Markov process,

$$\omega_{it} \;=\; g\left(\omega_{i,t-1}\right) + \xi_{it},$$

with ξ_{it} unforecastable at $t-1$, so that inputs chosen at $t-1$ are orthogonal to it and serve as instruments. The content of that assumption depends entirely on the length of the interval between $t-1$ and t.

In these data the interval between a firm's two observations ranges from **three to eighteen years**, with a median of six. Over six years — still more over thirteen, as in South Africa's 2007–2020 panel — the innovation ξ absorbs essentially all variation in productivity, $g(\cdot)$ carries almost no information, and lagged inputs retain no instrumental content. The estimator is not underpowered; its identifying assumption is violated by construction.

This deserves emphasis because it is a property of the data source, not of this sample. WBES panels are, by design, occasional re-interviews rather than annual accounts. Any study proposing a proxy-variable estimator on them faces the same obstacle.

7. First differences

With two observations per firm, the available estimator is the first difference,

$$\Delta y_i \;=\; \beta_L\,\Delta l_i + \beta_M\,\Delta m_i + \Delta\varepsilon_i,$$

which eliminates any firm-specific component of productivity constant between the two waves — the principal source of transmission bias.

Table 6 — Panel sample: pooled OLS versus first differences

	(1) Pooled OLS	(2) First difference
log employment / Δ log employment	0.202*	0.364**
	(0.118)	(0.152)
log materials / Δ log materials	0.748***	0.542***
	(0.081)	(0.103)
Implied median markup	2.31	1.68
Firm-years / firms	2,001	982
Economy clusters	14	14

Notes. Estimated on the 982 firms observed in two waves with production inputs complete in both. Column (1) pools both waves in levels; column (2) takes the within-firm difference, sweeping out permanent firm productivity. Standard errors clustered by economy. With 14 clusters these are optimistic; Table 7 reports wild bootstrap p-values. Markups use the panel-sample median materials share (0.31). *** p<0.01, ** p<0.05, * p<0.1.

Two things stand out. The materials elasticity falls substantially, from 0.748 to 0.542. And pooled OLS returns a labour elasticity of 0.202 — implausibly low for manufacturing — which first-differencing raises to 0.364, an internal sign that the pooled specification is badly contaminated.

Because the sample contains only 14 clusters, conventional cluster-robust inference is unreliable (Cameron, Gelbach and Miller, 2008). I therefore report a wild cluster bootstrap imposing the null, with Webb six-point weights, since Rademacher weights admit only 2^{14} distinct draws (MacKinnon and Webb, 2017).

Table 7 — Wild cluster bootstrap on the first-difference estimates

Coefficient	Estimate	CRV1 s.e.	CRV1 p	t	Bootstrap p (Webb)
Δ log materials	0.542	0.103	0.000	5.25	**0.0004**
Δ log employment	0.364	0.152	0.017	2.40	0.0766

Notes. Restricted wild cluster bootstrap imposing the null, 9,999 replications, Webb six-point weights over 14 economy clusters. Webb weights are used because Rademacher weights admit only 2^14 = 16,384 distinct draws at this cluster count (MacKinnon and Webb, 2017). The markup rests on the materials elasticity, which survives; the labour elasticity does not, so no claim about returns to scale is made.

The materials elasticity, on which the markup depends, survives decisively. The labour elasticity does not reach conventional significance, so **no claim about returns to scale is supportable** and none is made.

The first-difference markup of 1.68 sits within a few percent of the accounting benchmark of 1.81. It is tempting to read this as the structural and accounting approaches converging once the permanent component is removed. Section 8 shows that reading is wrong.

8. Attenuation and the within-country differencing-interval design

8.1 The problem with differencing

Differencing removes the fixed component of productivity, but it also removes much of the signal in the regressors while retaining the measurement error, amplifying attenuation bias. Griliches and Hausman (1986) analysed exactly this trade-off and made the observation on which this section rests: the degree of attenuation depends on the length of the difference taken, so differences of different lengths are informative about the measurement error itself. Where they considered annual panels with a choice of lag, the irregular spacing of firm surveys supplies the variation directly.

If the differencing interval is τ and true log inputs follow a process whose variance grows in τ while measurement error does not, the attenuation factor rises with τ: **longer intervals should yield larger estimated elasticities.**

8.2 The cross-country test is confounded

Table 8 — Materials elasticity by differencing interval, pooled across economies

Interval	Economies	Firms	θ_M	Implied median markup
3–4 years	6	261	0.272***	0.84
			(0.046)	
5–6 years	4	265	0.666***	2.06
			(0.175)	
7 years or more	6	456	0.692***	2.14
			(0.127)	
All	14	982	0.542***	1.68
			(0.103)	
Δm × interval (all firms)	14	982	+0.0497	p = 0.123
			(0.0322)	

Notes. First-difference specification of Table 6, column (2), estimated on the subsample named. **This split is confounded**: the differencing interval is a country attribute, so splitting on interval splits on country, and each subsample carries only three to six clusters. The interaction is not significant. Table 9 removes the confound. *** p<0.01, ** p<0.05, * p<0.1.

The ordering is as predicted, and the short-interval markup of 0.84 — below one, and so economically impossible — signals severe attenuation. But the test is invalid as it stands, for a reason that is easy to miss: **in a cross-country panel the differencing interval is a country attribute**. Splitting on interval is splitting on country. Each subsample carries three to six clusters, so no inference is possible, and the interaction between Δm and the interval is insignificant (coefficient 0.050, p = 0.123). Measured dispersion of the differenced regressor does not rise with the interval either (variance 7.63, 9.04, 6.62), contradicting the simplest version of the mechanism.

On this evidence attenuation is suggested and unproven.

8.3 Kenya breaks the confound

Kenya is surveyed in 2007, 2013, 2018 and 2025, and the World Bank's four-wave panel links firms across them. Because different firms are observed in different pairs of waves, the differencing interval varies *within a single country*: 46 firms at five years, 44 at six, 86 at seven, 22 at eleven, 21 at twelve and 5 at eighteen. Country, institutions, currency, survey administration and sampling frame are held fixed; only the interval varies. This is the design the rest of the paper depends on.

Table 9 — Materials elasticity by differencing interval, within Kenya

Interval	Firms	θ_M	Implied median markup
5 years	46	0.133	0.41
		(0.102)	
6 years	44	0.535***	1.65
		(0.104)	
7–12 years	129	0.683***	2.11
		(0.071)	
7 years or less	176	0.475***	1.47
		(0.070)	
More than 7 years	48	0.678***	2.10
		(0.098)	
All Kenya	224	0.501***	1.55
		(0.060)	
Δm × interval	224	**+0.0433****	**p = 0.0129**
		(0.0174)	

Notes. Kenya's four-wave panel (2007, 2013, 2018, 2025) is the one economy whose firms span many differencing intervals, so country, currency, institutions, sampling frame and survey administration are held fixed and only the interval varies. Heteroskedasticity-robust standard

errors. The interaction is the paper's central estimate: attenuation is identified within a single country. *** p<0.01, ** p<0.05, * p<0.1.

The gradient is monotone and steep. At a five-year interval the elasticity is 0.133 and statistically indistinguishable from zero, implying a markup of 0.41 — impossible, and therefore a direct demonstration that the estimate at short intervals is not measuring technology. At seven to twelve years it is 0.683. The interaction between Δm and the interval, estimated within Kenya, is

$$\hat{\beta}_{\Delta m \times \tau} \;=\; +0.0433 \quad (\text{s.e. } 0.0174), \qquad p = 0.013.$$

Attenuation is confirmed. The insignificant cross-country interaction of Section 8.2 was masked by the country confound.

Figure 2 shows the gradient with confidence intervals. The five-year estimate's interval contains zero; the seven-to-twelve-year estimate's does not come close. Reading the figure from left to right traces how much of the technology parameter measurement error destroys when the differencing interval is short.

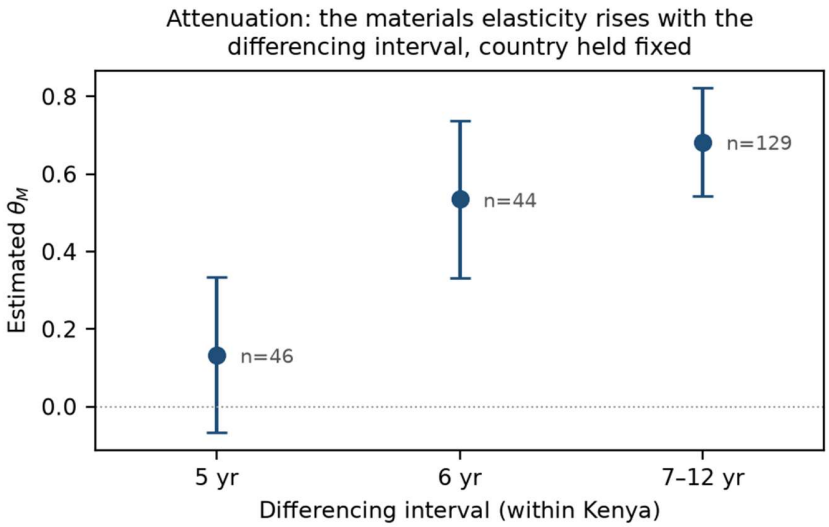


Figure 2 — Attenuation within Kenya. *Estimated materials elasticity by differencing interval, Kenya only, with 95 per cent confidence intervals from heteroskedasticity-robust standard errors. Country, currency, institutions, sampling frame and survey administration are held fixed; only the interval between the two observations varies.*

8.4 What this does to the first-difference estimate

The consequence is interpretive and it is the paper's central point. First-differencing does not *correct* ordinary least squares; it **exchanges an upward bias for a downward one**. OLS is contaminated by transmission bias, first differences by attenuation, and the apparent agreement between the first-difference markup (1.68) and the accounting benchmark (1.81) is best read as two opposing errors partially cancelling at this particular sample size and

composition. Indeed the agreement is unstable: at 758 firms the two figures were 1.80 and 1.77; adding Kenya moved them to 1.68 and 1.81. Estimates that move with sample composition in this way are not identifying a level.

Caveat. Longer intervals also involve more genuine input adjustment and stronger survivorship selection — a firm observed from 2007 to 2025 has survived eighteen years. The interval gradient is therefore not purely measurement error, though attenuation is the natural reading and the within-country design removes the principal alternative.

9. Partial identification

The two estimators bias in known and opposite directions. Ordinary least squares on the panel sample gives 0.748 and is biased upward; first differences give 0.542 and are biased downward. They bracket the parameter:

$$\theta_M \in [\,0.54,\ 0.75\,], \qquad \mu^{\text{median}} \in [\,1.68,\ 2.31\,].$$

Figure 3 displays the bracket. Neither endpoint is a point estimate of the technology parameter; each is an estimate whose bias has a known sign, and the interval between them is what the data support.

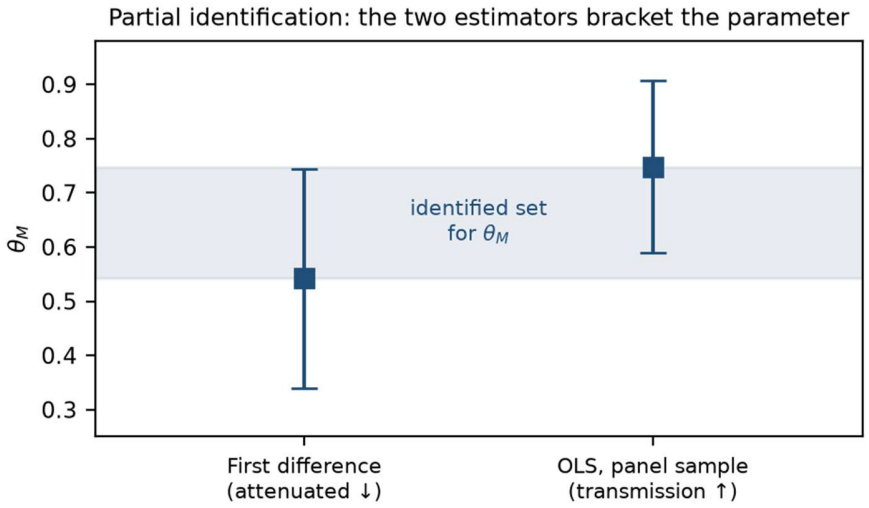


Figure 3 — Partial identification. *Point estimates with 95 per cent confidence intervals. Pooled OLS is biased upward by transmission; first differences are biased downward by attenuation, as established in Section 8. The shaded region is the resulting identified set for the materials elasticity.*

This is modest, and it is considerably more informative than the formal alternative. The Klepper–Leamer (1984) reverse-regression bound, computed on the same data by regressing the mismeasured input on output and the remaining regressor, admits elasticities up to **1.713** — an output elasticity of materials above one, which is economically meaningless. The

corresponding markup bound, [1.73, 5.46], is vacuous. When signal-to-noise is poor the reverse-regression bound carries no information, and the signed-bias bracket is the more useful object.

Three limitations bound the claim further. The estimates are **revenue-based**: without firm-level output prices, the estimated elasticity mixes technology with demand, a problem identified by Klette and Griliches (1996) and unresolved here. **Capital** is expenditure rather than a stock, so specifications including it answer a different question. And the accounting benchmarks impose constant returns and ignore the user cost of capital, making them a reference region rather than a target.

10. Discussion: how these results sit against the literature

10.1 The elasticities are ordinary; the markups are not

The estimated materials elasticity, 0.54 to 0.75, is unremarkable. Gross-output production functions routinely assign intermediate inputs an elasticity in the region of one half to two thirds, and the bracket here straddles that range. Nothing in the technology estimates signals a problem, which is precisely why the problem is easy to miss: **the pathology is invisible in the coefficient and appears only after division by the revenue share.** A referee inspecting Table 3's elasticities alone would pass them.

10.2 A direct contradiction with De Loecker and Warzynski, and what it means

The most useful comparison is also the most awkward. De Loecker and Warzynski (2012) report that markups "are estimated significantly higher when controlling for unobserved productivity" — controlling for productivity *raises* the estimated elasticity and hence the markup. Here, moving from pooled OLS to first differences, which is also a way of controlling for unobserved productivity, *lowers* the elasticity from 0.748 to 0.542 and the markup from 2.31 to 1.68. The sign is reversed.

The reversal is not a contradiction of their result; it is a consequence of the instrument available. De Loecker and Warzynski work with annual plant-level panels and a control function, which removes transmission bias while leaving the signal in the regressors intact. First-differencing over a six-year gap removes transmission bias *and* most of the signal, so attenuation dominates. The two studies are estimating the same parameter with tools whose biases point in opposite directions, and the direction of the correction reveals which force is binding in each setting.

This reading has a practical corollary. **Where a study reports that controlling for productivity lowers estimated markups, that is a signature of attenuation, not of successful bias correction** — and in short panels with long gaps it should be expected rather than celebrated. Section 8's within-country design is what allows this diagnosis to be made rather than assumed.

10.3 Comparison with markup levels in the literature

De Loecker, Eeckhout and Unger (2020) document aggregate United States markups rising from 21 per cent above marginal cost in 1980 to 61 per cent, that is from 1.21 to 1.61. Two features of their result matter for the comparison. First, their headline is a sales-weighted aggregate,

whereas the figures here are medians, and the two objects differ sharply when the distribution is skewed — as it is in Figure 1. Second, and more pointedly, they report that the **median United States markup is essentially unchanged** over the whole period, with the rise concentrated in the upper tail.

Set against that, a median markup of 2.11 for African manufacturing — higher than the United States aggregate at the peak of its documented rise, and far above the unchanged United States median — is not a plausible substantive finding. It is the reading that motivated this paper's diagnostic strategy, and the bracket of Section 9, [1.68, 2.31], does not resolve it: even the lower bound sits above the United States aggregate. The honest conclusion is not that African firms exercise more market power than American ones, but that the level is not recoverable from these data.

10.4 Productivity dispersion and the misallocation literature

The within-cell standard deviation of log total factor productivity reported in Section 3, 0.99, is large relative to the dispersion Hsieh and Klenow (2009) document for China and India, itself far above United States benchmarks. That would ordinarily be read as evidence of severe misallocation, consistent with Foster, Haltiwanger and Syverson (2008) on the role of reallocation in aggregate productivity growth.

Section 2.4 counsels caution. The measured dispersion is a residual, and the correlation matrix shows it is bound to the reported materials share at −0.50 by construction. Part of what presents as productivity dispersion is dispersion in a single survey answer. This does not overturn the misallocation reading — genuine dispersion in low-income manufacturing is well documented — but it means the *magnitude* inherits the same identification problem as the markup, and should carry the same caveat.

10.5 Relation to Griliches and Hausman

The design in Section 8 is an application of Griliches and Hausman (1986), who showed that in panel data with measurement error the attenuation of a differenced estimator depends on the length of the difference, and that differences of varying length are therefore informative about the measurement error itself. Their setting was an annual panel in which the researcher *chooses* the lag. The contribution here is to observe that repeated firm surveys supply that variation **without any choice being made**: waves are spaced irregularly, different firms are observed in different pairs of waves, and within a single country the differencing interval varies exogenously with survey timing.

That is why the result transfers. Most establishment surveys in low- and middle-income countries are irregular; the irregularity is normally treated as an inconvenience to be absorbed by fixed effects. It is instead the identifying variation for a measurement-error diagnostic, and it is available wherever a country has been surveyed at three or more unevenly spaced dates.

10.6 What the paper does not establish

Three limits should be stated plainly. The interval gradient within Kenya is consistent with attenuation but also with survivorship — firms observed over eighteen years are survivors — and with genuinely larger input adjustment over longer horizons; the design removes the country confound, not these. The estimates remain revenue-based, so the elasticity mixes technology with demand in the manner of Klette and Griliches (1996), and no firm-level price data exist in WBES to separate them. And the accounting benchmarks impose constant returns and ignore the user cost of capital, which is why they are used as a reference region rather than as a target the structural estimate should hit.

11. Conclusion

Markup and productivity *levels* are not identified from the World Bank Enterprise Surveys. The cross-sectional estimates that the data most readily support imply markups around two, which no plausible model of these markets sustains; the natural repair of broadening the input measure does not work, because the elasticity adjusts with the aggregate; the panel releases cannot support proxy-variable estimators, because their waves are years apart and the Markov assumption fails by construction; and first-differencing, the one estimator the two-wave structure permits, trades transmission bias for an attenuation bias that this paper measures directly. What remains identified is the *sign* of each bias and a bracket for the parameter.

For applied work the implication is concrete. Estimates of African firm markups or total factor productivity levels obtained from these surveys by cross-sectional methods should be read as upper bounds, and estimates from short-interval differences as lower bounds. The distance between them is not small.

The methodological contribution generalises beyond this application. Irregular wave spacing is usually treated as an inconvenience in repeated firm surveys. Section 8 shows it is an asset: when a single country is surveyed at several intervals, variation in the differencing interval identifies the attenuation bias directly, following the logic of Griliches and Hausman (1986) but using survey design rather than lag choice as the source of variation. Any repeated establishment survey with uneven waves — and most of them are uneven — admits the same diagnostic.

There is a constructive reading of the negative result. Knowing that a parameter is set- rather than point-identified, and knowing the direction of the error at each bound, is a usable finding. It is considerably more useful than a point estimate whose error has not been characterised.

11.1 Implications for the industrialisation debate

Return to the gap described in Section 1.1. Manufacturing's share of Sub-Saharan African GDP fell from 13.95 to 9.93 per cent between 2000 and 2024, against a Sustainable Development Goal target of doubling industry's share in least developed countries and an Agenda 2063 goal assessed at 17 per cent implementation. Explaining that shortfall requires choosing among firm-level diagnoses — low productivity, misallocation, entrenched market power — and each is

operationalised as a quantity this paper finds unidentified in the only data source that covers the region at firm level.

The implication is not that the diagnoses are wrong. It is that the *evidence base currently used to choose between them cannot do so on levels*. A finding that African manufacturing markups are high, offered as grounds for competition policy, may reflect a short differencing interval; a finding that productivity dispersion is extreme, offered as grounds for reallocation policy, partly reflects dispersion in one survey question. Policy conclusions that turn on magnitudes deserve the bracket, not the point estimate.

11.2 What would resolve it

Three things would. **Annual or biennial re-interviewing** in even a small panel of economies would restore the proxy-variable estimators, since the obstacle in Section 6.2 is wave spacing rather than sample size — a change in survey design, not in econometrics. **Firm-level output prices**, or a credible industry deflator at the firm level, would separate technology from demand and address the revenue-based critique. And a **capital stock question**, rather than the expenditure items WBES currently carries, would make the capital elasticity interpretable.

None of these is exotic; each is standard in the manufacturing censuses on which the canonical estimators were developed. The gap between what those estimators require and what low-income firm surveys supply is, in the end, the subject of this paper — and it is a gap in data collection, which is to say a gap that can be closed.

References

Ackerberg, D. A., Caves, K. and Frazer, G. (2015) 'Identification properties of recent production function estimators', *Econometrica*, 83(6), pp. 2411–2451. doi:10.3982/ECTA13408.

African Union Commission (2015) *Agenda 2063: the Africa we want*. Addis Ababa: African Union Commission.

Bond, S. and Söderbom, M. (2005) *Adjustment costs and the identification of Cobb Douglas production functions*. IFS Working Paper WP05/04. London: Institute for Fiscal Studies. doi:10.1920/wp.ifs.2005.0504.

Cameron, A. C., Gelbach, J. B. and Miller, D. L. (2008) 'Bootstrap-based improvements for inference with clustered errors', *Review of Economics and Statistics*, 90(3), pp. 414–427. doi:10.1162/rest.90.3.414.

De Loecker, J., Eeckhout, J. and Unger, G. (2020) 'The rise of market power and the macroeconomic implications', *Quarterly Journal of Economics*, 135(2), pp. 561–644. doi:10.1093/qje/qjz041.

De Loecker, J. and Warzynski, F. (2012) 'Markups and firm-level export status', *American Economic Review*, 102(6), pp. 2437–2471. doi:10.1257/aer.102.6.2437.

Foster, L., Haltiwanger, J. and Syverson, C. (2008) 'Reallocation, firm turnover, and efficiency: selection on productivity or profitability?', *American Economic Review*, 98(1), pp. 394–425. doi:10.1257/aer.98.1.394.

Griliches, Z. and Hausman, J. A. (1986) 'Errors in variables in panel data', *Journal of Econometrics*, 31(1), pp. 93–118. doi:10.1016/0304-4076(86)90058-8.

Griliches, Z. and Mairesse, J. (1998) 'Production functions: the search for identification', in Strøm, S. (ed.) *Econometrics and economic theory in the 20th century: the Ragnar Frisch centennial symposium*. Cambridge: Cambridge University Press, pp. 169–203. doi:10.1017/CCOL0521633230.006.

Hsieh, C.-T. and Klenow, P. J. (2009) 'Misallocation and manufacturing TFP in China and India', *Quarterly Journal of Economics*, 124(4), pp. 1403–1448. doi:10.1162/qjec.2009.124.4.1403.

Klepper, S. and Leamer, E. E. (1984) 'Consistent sets of estimates for regressions with errors in all variables', *Econometrica*, 52(1), pp. 163–183. doi:10.2307/1911466.

Klette, T. J. and Griliches, Z. (1996) 'The inconsistency of common scale estimators when output prices are unobserved and endogenous', *Journal of Applied Econometrics*, 11(4), pp. 343–361. doi:10.1002/(SICI)1099-1255(199607)11:4<343::AID-JAE404>3.0.CO;2-4.

Levinsohn, J. and Petrin, A. (2003) 'Estimating production functions using inputs to control for unobservables', *Review of Economic Studies*, 70(2), pp. 317–341. doi:10.1111/1467-937X.00246.

MacKinnon, J. G. and Webb, M. D. (2017) 'Wild bootstrap inference for wildly different cluster sizes', *Journal of Applied Econometrics*, 32(2), pp. 233–254. doi:10.1002/jae.2508.

Marschak, J. and Andrews, W. H. (1944) 'Random simultaneous equations and the theory of production', *Econometrica*, 12(3/4), pp. 143–205. doi:10.2307/1905432.

Mo Ibrahim Foundation (2024) *Assessing progress after the first decade (2014–2023) of Agenda 2063*. London: Mo Ibrahim Foundation.

Olley, G. S. and Pakes, A. (1996) 'The dynamics of productivity in the telecommunications equipment industry', *Econometrica*, 64(6), pp. 1263–1297. doi:10.2307/2171831.

Rodrik, D. (2016) 'Premature deindustrialization', *Journal of Economic Growth*, 21(1), pp. 1–33. doi:10.1007/s10887-015-9122-3.

United Nations (2015) *Transforming our world: the 2030 Agenda for Sustainable Development*. Resolution A/RES/70/1. New York: United Nations.

World Bank (2025) *World Development Indicators*. Series NV.IND.MANF.ZS, manufacturing value added (per cent of GDP). Washington, DC: World Bank.

Appendix A — Reproducibility

All results are produced by scripts in `WBES_ProdFunc/`, run in the order below.

Script	Produces
`build_pf_panel.py`	Harmonised cross-section, 20,047 firms (Section 2)
`estimate_pf_markups.py`	Cross-sectional elasticities and markups (Table 3)
`improve_markups.py`	Input-broadening test and accounting benchmarks (Tables 2, 3)
`extract_panels.py`	Extracts World Bank panel releases
`assess_panel_pf.py`	Panel linkage and completeness audit (Section 6.1)
`fd_estimate.py`	First-difference estimates (Table 6)
`fd_wildboot.py`	Wild cluster bootstrap (Table 7)
`attenuation_test.py`	Cross-country interval split and Klepper–Leamer bound (Table 8)
`attenuation_within_kenya.py`	Within-Kenya design (Table 9, Figure 2)
`build_corr_figs.py`	Correlation matrix (Table 2) and Figures 1–3
`out/audit_refs.py`	Crossref verification of every reference in the list

`acf_estimate.py` implements the Ackerberg–Caves–Frazer estimator and is retained for completeness; Section 6.2 explains why it is not run on these intervals.

Three data traps are handled in code and noted here for anyone rebuilding the pipeline: the ISIC revision change that makes two-digit codes non-comparable across survey generations (Section 2.2); the panel key named for the earlier wave, and the file carrying both a dead and a live key (Section 6.1); and Kenya's panel release, whose variable labels are Latin-1 encoded and require an explicit encoding argument, together with the fact that its three nested releases must not be stacked.