The Effects of Tax Enforcement on Average Firm Size and Aggregate Productivity: Appendix

A Cross-Country Relationship between Informality and TFP

To use measures of informality that are consistent across countries and over time, I rely on the database in Elgin et al. (2021), which provides estimates of informality rates for more than 160 countries over the period 1990-2019. Elgin et al. (2021) introduce two time series who stem from different methods and assumptions. The first method, whose associated database I employ in the paper for cross-country comparisons, is based on a *dynamic general equilibrium* (DGE) model developed by Elgin and Öztunali (2012). In the model, a representative household decides how much labor to allocate to the formal and the informal sector. The *Multiple Indicators Multiple Causes* (MIMIC) in Schneider, Buehn, and Montenegro (2010) constitutes the source of the second series. The method consists of a simultaneous specification of a factor (measurement) model and a structural model. The main idea is to retrieve an unobserved variable (share of informal economy) using structural equations and the sample covariances between observed variables.

Table A1 shows the world averages of informal share of GDP and TFP from 1990 to 2018. Using the DGE based measure of informality as a benchmark (column 1), the table highlights that the informal share of GDP has decreased by about 8 percentage points from 34.8 to 26.8 in the period considered.[1] In the same period, TFP has increased by about 6 percentage points on average (column 4). Column (3) displays the average estimate of TFP as a percentage of US TFP. This is a measure used for comparisons of TFP across

1. Informality has also decreased according to the MIMIC measure of informality, but by about 3 percentage points (column 2).

countries at a given point in time.

Figure A1 depicts the cross-country correlation between informality and TFP. The left panel shows the DGE-based measure, while the right panel displays the MIMIC-based measure. TFP is measured as a percentage of US TFP. The two panels point out a negative correlation between TFP on the one hand and both measures of informality on the other hand.

Table A1: Informality and TFP. Simple Averages.

Year	% Informal GDP (DGE) (1)	% Informal GDP (MIMIC) (2)	TFP (US=100) (3)	TFP (2017=100) (4)
1990	34.8	.	72.4	94.4
1991	34.7	.	71.0	93.1
1992	34.3	.	69.9	93.8
1993	34.3	34.7	68.7	93.6
1994	34.1	34.6	65.4	88.8
1995	34.0	34.5	65.5	88.8
1996	33.8	34.3	65.7	89.5
1997	33.6	34.2	65.2	90.4
1998	33.4	34.2	63.5	90.3
1999	33.2	34.2	63.3	90.3
2000	33.1	34.0	65.3	91.1
2001	32.9	34.0	65.0	91.6
2002	32.7	34.1	65.3	92.6
2003	32.5	34.0	64.8	93.7
2004	32.3	33.7	65.1	96.2
2005	32.2	33.6	67.2	97.5
2006	31.9	33.2	68.0	99.2
2007	31.6	32.9	68.4	100.4
2008	31.3	32.8	68.0	99.9
2009	30.9	33.4	64.8	97.8
2010	30.7	33.0	64.7	99.3
2011	30.4	32.9	66.8	100.1
2012	30.0	32.8	67.4	100.4
2013	29.7	32.7	66.9	100.4
2014	29.4	32.6	65.9	100.3
2015	29.1	32.5	63.8	99.7
2016	28.9	32.4	63.7	99.6
2017	28.7	32.1	64.4	100.0
2018	26.8	31.9	64.0	99.9

World averages. Sources: Informality data are from Elgin et al. (2021). TFP data are from Feenstra, Inklaar, and Timmer (2015).

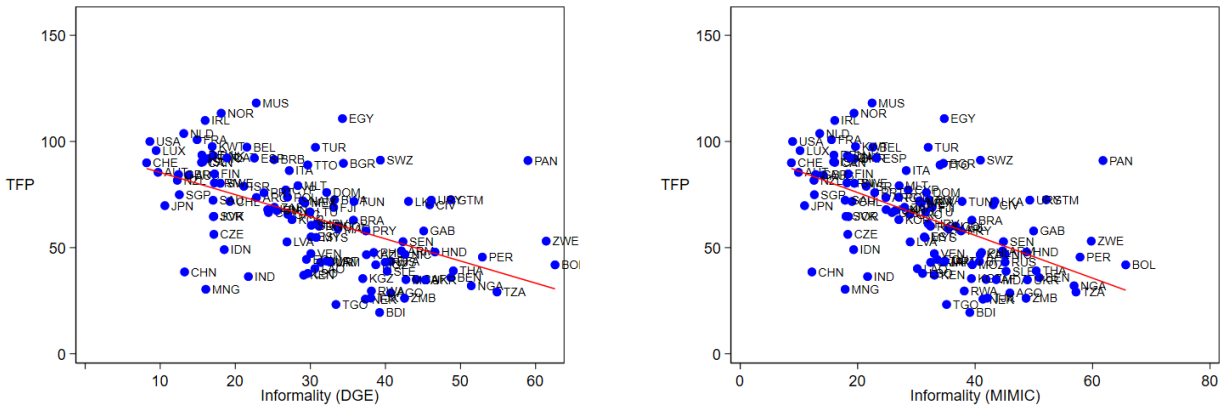
To further assess the relationship between informality and TFP, I perform a series of cross-country regressions. First, I run an OLS regression according to the following specification:

$$TFP_{it} = \alpha_0 + \alpha_1 INF_{it} + \gamma GDP_{it} + Year_t + \epsilon_{it}, \tag{A1}$$

where TFP_{it} is TFP in country i in year t (relative to US TFP in year t), INF_{it} is the share

Figure A1: Correlation between Informality and TFP across Countries.

(a) Informality (DGE) and TFP (US=100). (b) Informality (MIMIC) and TFP (US=100).



Data sources: Elgin et al. (2021) for informality, Feenstra, Inklaar, and Timmer (2015) for TFP. Each data point represents the country's averages of informality and TFP over the period between 1990 and 2018.

of informal output, GDP_{it} is GDP per capita, $Year_t$ denotes year fixed effects, and ϵ_{it} is the error term. α_1 is the coefficient of interest in the relationship between informality and TFP. Since the two series of informality are highly correlated, I adopt the DGE-based measure, which is the one used in the quantitative experiments in Section **??**.

Second, I exploit the time series dimension of the data to perform a panel regression:

$$TFP_{it} = \alpha_0 + \alpha_1 INF_{it} + \gamma GDP_{it} + c_i + \epsilon_{it}, \tag{A2}$$

where c_i denotes country-specific fixed effects. Since this type of panel regression captures the effect within a country, the correct measure of TFP to use is the one relative to a base year (2017) for that country, that is, the series whose cross-country averages are shown in column (4) of Table A1.

Table A2 shows the results of the regressions described above. Column (1) displays the coefficient of the OLS specification without GDP per capita, while column (2) includes it. The inclusion of GDP per capita reduces the magnitude of the coefficient of interest, which remains negative and statistically significant. A coefficient of -0.26 can be interpreted in the following way: a decrease in informality by one percentage point is associated with an increase in TFP (relative to US TFP) by 0.26 percentage points. Column (3) and column

Table A2: Relationship between TFP and Informality.

	(1) OLS TFP (US=100)	(2) OLS TFP (US=100)	(3) Panel TFP (2017=1)	(4) Panel TFP (2007=1)
Informality (DGE)	-1.026***	-0.261***	-0.879***	-0.852**
	(0.0313)	(0.0344)	(0.321)	(0.339)
GDP per capita (th)		0.828***		0.0271
		(0.0237)		(0.227)
Year fixed effects	YES	YES	NO	NO
Country fixed effects	NO	NO	YES	YES
Observations	3198	3198	3198	3198

Standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

(4) display the results of the panel regressions excluding and including GDP per capita as control, respectively. The coefficients on informality are negative and statistically significant. In this case, a coefficient of -0.85 means that a reduction in informality by one percentage point is associated with an increase in TFP by 0.85 percentage points.

B Brazilian Data

ECINF.[2] *Pesquisa de Economia Informal Urbana* (ECINF) is a survey conducted by *Instituto Brasileiro de Geografia e Estatística* (IBGE), the Brazilian Bureau of Statistics. It was conducted in 1997 and 2003 to collect information about the informal sector. The survey is nationwide representative for small non-agricultural businesses with a maximum of 5 employees. Owners are classified as informal if they do not possess a tax identification number (*Cadastro Nacional de Pessoa Juridica*, CNPJ).[3] By matching owners and businesses with employees, it is possible to obtain the number of employees for each business.

From the original dataset, which is publicly available, I exclude the following observations:

2. Appendix B has strongly benefited from the Data Zoom (2023) project and the Ulyssea (2018) replication package.

3. Ulyssea (2018) points out that strict confidentiality clauses and IBGE's reputation induce respondents to report fairly accurately.

1. Owners who operate in the agricultural and construction sectors.
2. Owners who lack a facility exclusively dedicated to the business outside their house.
3. Owners who have another job.
4. Businesses who have no owners or more than 4 owners.
5. Businesses who have more than 7 employees.

2 and 3 directly address the concern that the data might measure home production rather than entrepreneurship (as pointed out by Erosa, Fuster, and Martinez (2023)).
The final sample from which I compute moments used in the calibration contains about 30,000 firms. Table B3 summarizes sector composition and size distribution by formality status.

Table B3: ECINF. Sector and Size Statistics.

	Formal Shares	Informal Shares	Total Shares
Industry			
Manufacturing	8.73	11.24	10.82
Retail	51.08	37.66	39.93
Services	40.19	51.09	49.25
Number of Employees			
1	28.26	79.13	70.53
2	27.38	14.23	16.45
3	16.51	3.72	5.89
4	11.93	1.73	3.46
5	8.81	0.81	2.16
6	4.58	0.30	1.03
7	2.53	0.07	0.49
Observations (#)	5,261	24,924	30,185

Source: Own calculation from ECINF (2003). Note that the shares reflect the raw unweighted sample.

References

Data Zoom. 2023. "Data Zoom: Simplifying Access To Brazilian Microdata." Available at https://www.econ.puc-rio.br/datazoom/english/index.html. https://www.econ.puc-rio.br/datazoom/english/index.html.

Elgin, Ceyhun, M. Ayhan Kose, Franziska Ohnsorge, and Shu Yu. 2021. "Understanding Informality." *Working paper.*

Elgin, Ceyhun, and Oguz Öztunali. 2012. "Shadow economies around the world: Model based estimates." *Working Papers 2012/05. Bogazici University.*

Erosa, Andrés, Luisa Fuster, and Tomás R. Martinez. 2023. "Public Financing with Financial Frictions and Underground Economy." *Journal of Monetary Economics* 135:20–36.

Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer. 2015. "The Next Generation of the Penn World Table." *American Economic Review* 105 (10): 3150–3182.

Schneider, Friedrich, Andreas Buehn, and Claudio Montenegro. 2010. "New Estimates for the Shadow Economies all over the World." *International Economic Journal* 24 (4): 443–461.

Ulyssea, Gabriel. 2018. "Firms, Informality, and Development: Theory and Evidence form Brazil." *American Economic Review* 108 (8): 2015–2047.