CESifo WORKING PAPERS

10461
2023

May 2023

# The Impact of Inequality on the Informal Economy in Latin America and the Caribbean with a MIMIC Model

*Marcos Takanohashi, Marcel Ribeiro, Friedrich Schneider*

CESifo

**Impressum:**

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10461*

# The Impact of Inequality on the Informal Economy in Latin America and the Caribbean with a MIMIC Model


## Abstract

Vast literature is available covering main Informal Economy (IE) causes and consequences for Latin America and Caribbean (LAC), but its size estimation has been mainly limited to worldwide models applied to the region. This paper proposes a MIMIC Base Model using a data set composed by 41 countries in LAC, in which both inequality and total factor productivity are introduced on top of traditional variables. Modelling results on IE estimates for LAC countries that are at par with the literature. When compared with a model using data from 188 countries, inequality has an impact ten-fold higher in the level of informality. Results suggest the importance of tailored data set selection when modelling IE with MIMIC.


JEL-Codes: D780, E260, H260, H320, O170, O540.

Keywords: Informal Economy (IE). MIMIC (Multiple Indicators Multiple Causes), shadow economy, underground economy, non-observed economy, inequality, informality, tax burden, quality of state institutions, regulations, governance, GINI coefficient, TFP, Latin America and Caribbean (LAC).

*Marcos Takanohashi*
*São Paulo School of Economics, FGV*
*São Paulo / Brazil*
*marcos.takanohashi@fgv.edu.br*

*Marcel Ribeiro*
*São Paulo School of Economics, FGV*
*São Paulo / Brazil*
*marcel.ribeiro@fgv.br*

*Friedrich Schneider*
*Johannes Kepler University*
*Linz / Austria*
*Friedrich.Schneider@jku.at*

May 23, 2023

# 1 INTRODUCTION

Estimates of the size of Informal Economy (IE) in Latin American and Caribbean (LAC, henceforth) countries are provided by the World Bank and several authors using different modelling tools. In general, they are built with a data set composed of a cross panel of worldwide countries, such as the World Bank 160-country sample. These estimates assume that causes have homogeneous effects on IE and hence some standard variables play essential roles in determining IE worldwide. However, some other variables might be more prominent in certain regions than others, eventually having higher relevance in some geographies than in others, producing heterogeneous effects.

While estimating heterogeneous effects on latent variable models is challenging, a straightforward alternative is to evaluate if the parameter estimates for regional MIMIC specifications are consistent with the worldwide equivalent. By considering this regional approach, some insights about the drivers of informality can be shed light. We apply such insight in the context of the LAC countries. Our results suggest that worldwide estimations overlook an essential element of those countries: widespread and persistent income inequality. Specifically, the GINI index coefficient estimate is roughly ten times higher in LAC countries' specification than in the worldwide specification. This result has key implications for policy recommendations in those countries and the methodology of estimating informality.

Using the framework proposed by Schneider, Buehn & Montenegro (2010) and leveraging on the analysis presented by Matos & Veiga (2019), the present article has the objective of further contributing to the long path of understanding IE by proposing a MIMIC (Multiple Indicators Multiple Causes) model tailored to LAC, using regional data, and leveraging on factors and variables that are relevant to our reality. With the proposed model we then estimate IE values for countries in LAC from 1996 to 2019, as percentage points of GDP.

The fine line separating legal from illegal shadow economic activities is a matter of interpretation, not classification, what further complicates the matter. One is genuinely purchasing pay TV services from a local Service Provider in a *favela* in Rio, but this company is just privately broadcasting signals illegally acquired without paying royalties or rights; the company itself is most likely not incorporated and it does not collect any taxes. The consumer is getting a service level, with local support, but he/she's part of a much broader crime ecosystem that's part of IE in LAC. Same happens with electricity, gas, water, phone, and broadband Internet, whereas in *favelas, villas* or slums all over LAC countries. These informal arrangements happen in a multitude of situations, from someone extending a wire to get electricity from a neighbor to an organized illegal enterprise such as a militia in Rio de Janeiro controlling all broadband internet distribution in a *favela*.

Literature focused on LAC has been unclear on what to measure: sometimes it focuses on the informal sector excluding illegal activities, others it tries to capture underground activities that are not 100% illegal as per the taxonomy presented on OECD (2002). Moreover, in the whole region we observe the phenomena of semi-formality (NOGUEIRA, 2016), where formal and informal relations blend into a new semi-formal category. Researchers as Chen (2012) have placed focus on extending the framework of the non-observed economies and its relations with the formal sector, by proposing a classification of different informal sectors and its historical evolution. Dell'Anno (2022) presents a comprehensive survey of the existing literature, covering the recent effort to provide a common taxonomy for it.

Since informal, semi-formal and illegal activities blend into intertwined relations, we focus on measuring the complete non-observed economy, including informal, underground, and illegal activities not captured by formal GDP measures. Within ILO (2021) proposed taxonomy, we aim to measure all informal productive activities, as well as illicit production and underdeclared market value transactions of ordinary goods and services. Throughout this article we will use terms such as informal, shadow, underground to reflect the non-observed economy as defined hereinbefore.

Better understanding of shadow economy in the region started by Soto (1989) with the framing of the problem, followed by numerous articles describing informality aspects in different countries in LAC in the 1980s and 1990s. Social studies bring to the surface aspects related to the colonization of our countries, and the clash of social classes where establishment was created upon privilege building at higher ranks. Much has happened since then, and while societies and institutions have evolved, some embedded aspects remain and are probably reflected in the way people face labor and governments. Gutiérrez-Romero (2021) provides a comprehensive discussion relating the persistence of inequality as a strong contributor to the size of IE in a country.

Late 1990s and early 2000s researchers started to tackle the main causes and consequences of IE (TANZI, 1999) on frameworks attempting to measure its size either via microeconomics models deriving from the seminal work by Rauch (1991) or via macroeconomics models like the MIMIC methodology proposed by Jöreskog & Goldberger (1975) and leveraged by Schneider & Enste (2000). Other methods are used, and Medina & Schneider (2019) present a vast array of tools to size IE, from surveys to macro and micro models.

This article uses MIMIC to estimate the size of IE in countries in LAC. Matos & Veiga (2019) provide a similar approach to a set of 18 LAC countries; here the data set comprises of 41 countries during the period from 1996 to 2019, and our analysis is focused on comparing estimated parameters using regional and worldwide data sets, aiming to identify homogeneous and heterogeneous effects on the different specifications.

With MIMIC, structured equation models and measurement models are used to estimate IE as a latent variable of certain causes and certain indicators. Causes are factors that have direct impact on the shadow economy, while indicators are variables that are impacted by the shadow economy. By observing causes and indicators, MIMIC allows us to estimate the latent IE.

We propose special focus on the particularities of our region, such as on the selection of data, on proposing new variables, and on the interpretation of results. Two proposed causal variables present robust results in our model: GINI as a measure of inequality, and TFP as a measure of total factor productivity.

Inequality has always been a key factor observed in LAC societies. The main intention to add inequality to our model is to get full advantage of it once several common effects can be offset such as cultural and colonization aspects. Using it with regional data seemed different to using it with worldwide heterogeneous data, and results show inequality gets more relevant on causing IE compared to other studies.

Total factor productivity (TFP) is proposed with the intention to identify a channel to firms' formalization. Obtained results show however that TFP is potentially a channel for both firms' formalization as well as informal labor utilization.

Focusing our analysis to regional specifications may contribute to the literature, in the sense of putting into perspective the heterogeneous effects some aspects such as inequality have on informality, shedding light for tailored policy recommendations that can contribute to long term IE improvements on LAC.

## 2 DATA AND ECONOMETRICS

The use of MIMIC with structured equations and Informal Economy (IE) as a latent variable is per se a challenge. It was probably more difficult early 2000s when several studies flourished proposing interesting models with limited data availability. More recent studies take advantage of the learning obtained in past years, as well as of more polished data availability, data manipulation techniques and computer processing power, to make it a more reliable methodology. It's safe to say that structured equations with latent variables are being much more used in different science areas, with IE being a very active one.

MIMIC consists of two parts, a Structural Model, and a Measurement Model. Figure 1 provides a graphical representation of the MIMIC model, with q causal variables and p indicator variables:

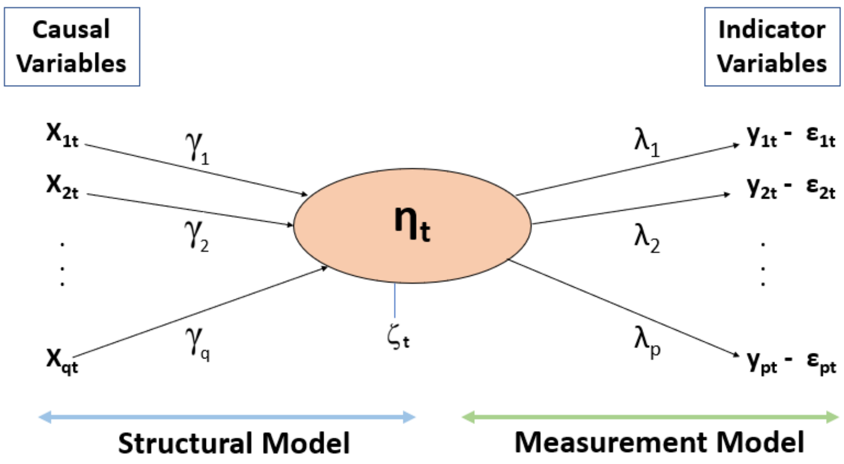


Figure 1 – MIMIC model - general structure

The Structural Model linearly determines the latent variable $\eta_t$, which represents the size of IE, from a vector of q observable exogenous causes $X_t' = (X_{1t}, X_{2t}, ..., X_{qt})$, subject to disturbances $\zeta_t$ representing the unexplained component. $\gamma = (\gamma_1, \gamma_2, ..., \gamma_q)$ contains the coefficients that weight each causal variable on the estimation of the latent variable. Usage of MIMIC aims to estimate $\gamma$ and allow us to size IE from some observed causal variables. The structural equation is as follows:

$$\eta_t = \gamma * X_t + \zeta_t$$

The Measurement Model projects the effects of the latent variable $\eta_t$ on a vector of p observed variables $y_t' = (y_{1t}, y_{2t}, ..., y_{pt})$, subject to disturbances $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t}, ..., \epsilon_{pt})$ representing the error term. $\lambda = (\lambda_1, \lambda_2, ..., \lambda_p)$ represents the magnitude of the change on the corresponding indicator variable for a unit change in the latent variable $\eta_t$. The measurement equation is as follows:

$$y_t = \lambda * \eta_t + \epsilon_t$$

MIMIC then estimates the latent variable considering both Structural and Measurement models, using maximum likelihood to converge a solution. The resulting latent variable $\eta_t = (\eta_{1t}, \eta_{2t}, ..., \eta_{jt})$ provides IE estimates for j countries considered in the data set. Actually, $\eta_t$ provides variations year-on-year of IE for these countries.

Breusch (2005) presented significant criticism on MIMIC modelling and how data was manipulated in previous works, in terms of using a combination of time series from different sources and losing its traceability. Dell'Anno & Schneider (2006) presented responses to Breusch's critique by addressing each topic and providing a solution path that assures MIMIC as a valid modelling tool. We managed data observing Breusch's critique and addressing them in line with Dell'Anno's responses.

MIMIC can eventually be considered a big data algorithm that leverages on data richness to obtain results for non-measured factors; as such, data richness produces improved results. One area of focus to produce our model was the selection of available data for LAC, and the variables to our MIMIC model. Taking advantage of broader data availability, we have restricted data manipulation as much as we could. Most of the time series used as observed variables are taken from sources like the World Bank, IMF (International Monetary Fund), ILO (International Labor Organization), United Nations, Penn World Table, and the Heritage Foundation. For region-specific focus, we used data from LatinoBarometro and processed by the World Bank in their Worldwide Governance Indicators.

The World Bank (OHNSORGE; YU, 2022) provides estimates of the informal economies to a broad list of countries worldwide, using both Multiple Indicators Multiple Causes (MIMIC) and Dynamic General Equilibrium (DGE) models, with compatible results. These are the two main econometric modelling techniques being used today. Estimating the size of IE, the Shadow Economy, The Underground Economy, or the Non-Observed Economy is a significant challenge, as measurement is always done indirect, whereas via sample surveys, whereas by means of estimations with MIMIC or DGE models.

Table 1 presents the observed variables used in our MIMIC Base Model, both causal and indicators, with the reference to the data sources, and expected estimator signs, whereas positive or negative, as per existing literature and economic intuition.

Dell'Anno & Schneider (2009), Trebicka (2014) and Schneider & Buehn (2013) provide a comprehensive description of the MIMIC technique, including the structural model examining the relationships between the causal variables and the latent variable IE, and the measurement model (or factor model) linking indicator variables to IE. As IE is latent and non-observed, MIMIC minimizes the difference between the covariance matrix of sample data of observed variables and the model predicted covariance matrix. On doing that, the identification of the model requires the normalization of one of the elements of the set of estimators of indicator variables. In our Base Model we implement the normalization on the Services GDP indicator variable, i.e., $\lambda_{ServicesGDP} = 1$.

Our proposed Base Model uses the following causal variables:

1. **Taxes** – the literature indicates that the higher the tax burden, the higher IE (Informal Economy), ceteris paribus; expected sign is positive. Taxes directly impact labor-leisure choices, and the higher the taxes, the less one must spend. To better reflect the tax burden in LAC, we use three proxy variables traditionally used in the literature: Direct Taxes over Total Taxes, Size of Government and Total Taxes over GDP.

2. **Regulations & Governance** – this is probably the most challenging area to select variables. Even though there are several reliable sources of Governance key indicators, such as World Bank's Worldwide Governance Indicators, Freedom House, Economist Intelligence Unit, Vanderbilt's Americas Barometer, World Justice Project, the survey data consolidation mixes up different perspectives to build common indicators such

| | | Variable Name | Variable Description | Source of data | Data Series | Expected Sign |
|---|---|---|---|---|---|---|
| | Indicator Variables | ServicesGDP | Services component of country GDP | World Bank | NV.SRV.TOTL.ZS | |
| | | LaborForceGrowth | Labor Force Growth Rate | ILO | EAP_2EAP_SEX_AGE_NB_A | Negative |
| | | Currency | M0 over M1 - currency in circulation over monetary base | IMF | Monetary Financial Statistics SRF (Standardized Report Forms) and non-SRF reports | Positive |
| | | LaborForceParticipation | Labor Force Participation Rate | ILO | ILO_EAP_2WAP_SEX_AGE_RT_A | Negative |
| | | EnergyConsumptionGrowth | Energy Consumption Growth in Joules | EIA - US Energy Information Administration | https://www.eia.gov/totalenergy/data/annual/ | Negative |
| Causal Variables | Taxes | DirectTaxesOverTotalTaxes | Proportion of direct taxes (over income, profits and capital gains) over total taxes | IMF, World Bank, OECD | OECD Global Revenue Statistics Database | Positive |
| | | SizeOfGovernment | Government expenses over GDP | United Nations | Government component of GDP breakdown by expenditure | Positive |
| | | TaxOverGDP | Total taxes over GDP | World Bank | GC.TAX.TOTL.GD.ZS | Positive |
| | Regulatory & Governance | TradeFreedom | Indicates the level of market openness to overseas trade | Heritage Foundation | https://www.heritage.org/index/explore?view=by-region-country-year&u=637977061980012970 | Negative |
| | | BusinessFreedom | Indicates the regulatory efficiency to do business in a country | | | Negative |
| | | EconomicFreedom | Compounded index for all Heritage Foundation metrics | | | Negative |
| | | LaborFreedom | Indicates the regulatory efficiency of the local labor regulation | | | Negative |
| | Public Sector | LB_GovernmentEffectiveness | Represents the Government effectiveness and the trust people have in institutions | LatinoBarometro | https://info.worldbank.org/governance/wgi/Home/downLoadFile?fileName=LBO.xlsx | Negative |
| | | LB_ControlCorruption | Reflects the control of corruption at different levels in public administration and country entities | | | Negative |
| | Offical Economy | Unemployment | Unemployment rate - ILO modelling | World Bank | SL.UEM.TOTL.ZS | Positive |
| | | GDPpercapitaPPPgrowth | GDP growth rate per capita PPP | | NY.GDP.PCAP.PP.CD | Negative |
| | New proposed variables | GINI | Indicator for inequality | World Bank | SI.POV.GINI | Positive |
| | | PWT_ctfp_h | Total Factor Productivity PPP per country relative to the USA | PWT 10.0 | "ctfp" data series | Positive or Negative |

Table 1 – MIMIC Base Model - description of variables

as, for instance, Government Effectiveness. As proposed by Schneider, Buehn & Montenegro (2010) we use a set of variables from the Heritage Foundation to provide our framework with key governance indicators: Trade Freedom, Business Freedom, Labor Freedom, and Economic Freedom. The higher these indicators are, the higher the intensity of regulation on the specific field under analysis, whether labor, trade, business, or economy. In general, regulations result on an increase of labor costs in the formal economy, consisting of another incentive to the informal sector. The higher the regulation, the lower IE, ceteris paribus; negative sign expected.

3. **Public Sector** – we propose to use some variables from LatinoBarometro (LB), a private non-profit organization incorporated in Chile that provides political and social data based on regular surveys conducted since 1995. We use two variables, Government Effectiveness and Control of Corruption. When running our Base Model to other geographies, we then use WGI's Government Effectiveness and Control of Corruption time series in lieu of LatinoBarometro. As with governance, an increase on these indicators reflects better framework. Ceteris paribus, the higher these indicators, the lower IE; expected sign is negative.

4. **Official Economy** – as proposed in the literature, we use Unemployment and GDP per Capita growth. Expected sign for Unemployment is positive, as IE tends to grow as unemployment rises. Sign for GDP per Capita growth is negative, as IE tends to reduce with economic growth.

5. **New proposed causal variables** – We have decided to propose two variables that

maintain close ties to economic development in LAC and that, traditionally, have been statistically relevant in studies and analysis made in other economic fields:

- *A measure of inequality* – we suggest using **GINI** from the World Bank. Inequality is a key factor in LAC that impacts and limits economic development, and we expect higher GINI to cause growth on the informal sector, ceteris paribus; positive sign expected.
- *A measure of total factor productivity* – we suggest using the PWT 10.0 variable PWT-ctfp-h, which provides the level of **TFP** PPP per country relative to the USA. PWT's ctfp data series is broadly used to provide a balanced comparison on the total factor productivity across countries. The expected sign is negative.

Our model uses five indicator variables: Services GDP, Labor Force Growth Rate, Currency, Labor Force Participation Rate and Energy Consumption Growth. In LAC, informal labor is usually more tied to the services segment, and as such we've decided to use Services GDP instead of GDP.

Our MIMIC model is of type 13-1-5, with 13 causal variables, one latent variable (IE, the Informal Economy estimate) and 5 indicator variables. The literature, in general, tend to use 2 or 3 indicator variables. As econometric software has evolved, and data processing is less of a challenge nowadays, we have decided to use more indicator variables to take advantage of the MIMIC econometrics leveraging on further computing processing to refine likelihood calculations.

Figure 2 provides a graphical representation of our proposed Base Model. The plus and minus signs preceding the name of causal and indicator variables are the expected estimator signs from the modelling.

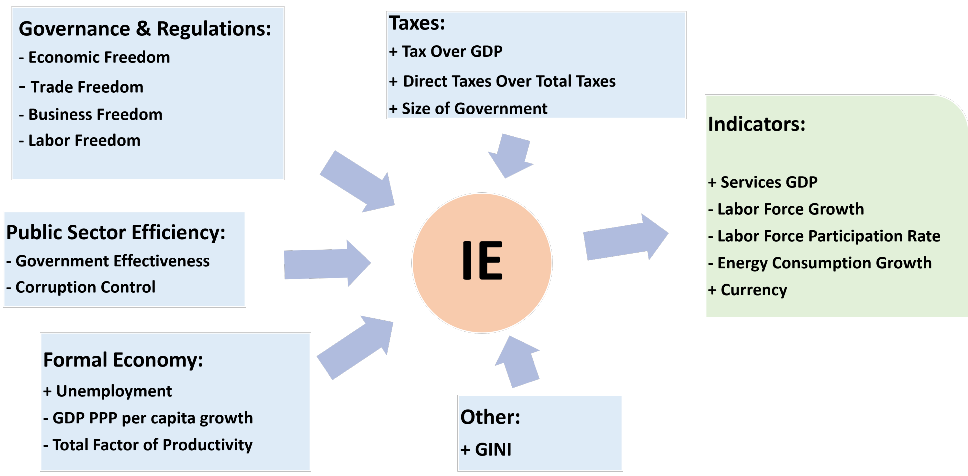


Figure 2 – Proposed MIMIC model with causal and indicator variables.

The choice of variables is key on MIMIC models, and therefore we presented a detailed review in this section. When trying to replicate the econometrics done by other authors such as the works described in the previous section, we have faced numerous challenges to replicate the construction of their databases.

Certain authors decide to work with the level while others decide to differentiate certain time series data to extract their trends (unit root). Breusch (2005) provides a thorough discussion of what's the right path, and we decide to follow his recommendation by working with the level and avoid any manipulation of time series data used in our work. Rationale for this is the construction of MIMIC structured equations as a set of cross section panels with a maximum likelihood approach to estimate the latent variables, rather than traditional time series manipulation where unit roots need to be extracted.

We implement our MIMIC model using the "lavaan" library on R Studio[1].We use a database of 41 countries in LAC, from 1995 to 2019 as our main data set. This period is used throughout the different models and geographies run. Now on we refer to this data set as Base Sample.

In Appendix I we present the statistics for causal and indicator variables used in our model (Table 5), as well as results of experiments done with our proposed Base Model to different sets of countries (data set samples). Appendix II presents an Alternative Base Model, using MIMIC 14-1-4 instead of the MIMIC 13-1-5 from the Base Model, and main results are presented in Section 3.

Finally, the current MIMIC modelling provides relative estimates, not absolute; we get year-on-year increase/decrease of IE estimated size using lavaan's "lavPredict" function for each country. "lavPredict" considers both Structural and Measurement models to come up with year-on-year increase-decrease estimates.

Then we process a calibration (or benchmark) procedure to come up with absolute values. For such, an anchor value is needed, and we use guidelines provided by Schneider, Buehn & Montenegro (2010), using year 2000 as an exogenous anchor reference for our estimates; such reference is calculated by the authors with the Currency Demand method. The following formula is used to calculate absolute IE estimates:

$$\hat{\eta_t} = (\tilde{\eta_t}/\tilde{\eta_{2000}}) * \eta^*_{2000}$$

Results from "lavPredict" are indicated by $\tilde{\eta_t}$, with the relative increase/decrease estimates of our latent variable, and $\tilde{\eta_{2000}}$ represents relative IE in year 2000. Absolute IE size is represented by $\hat{\eta_t}$, and $\eta^*_{2000}$ represents absolute IE in year 2000, with exogenous values from Schneider (2007).

# 3 MAIN RESULTS

Table 2 shows the different data set Samples running our proposed Base Model, with the resulting estimators. Our Base Sample is composed by a cross panel of 41 LAC countries, with data from 1995 to 2019; Sample 2 is the same cross panel, including data from 2020 to understand the model behavior with the COVID impact; Sample 3 contains the cross panel of countries considered by Schneider, Buehn & Montenegro (2010), to compare the behavior of our model with data coming from a cross panel of worldwide countries. Finally, on the last columns of this table we included the results of our Base Model excluding GINI.

GINI is statistically relevant on all models, with estimators 10x higher on Samples with regional data (Base Sample and Sample 2) compared to global data (Sample 3). Table 2

[1] Plenty of literature explain the MIMIC model, and the usage of lavaan to implement MIMIC over structured equations. We decided to implement our models in lavaan emulating Mplus (CHANG *et al.*, 2020), which is actually very close to lavaan's implementation of MIMIC. Additionally, as our data series contain several missing values, we used the option missing="ML", asking lavaan to estimate the full information maximum likelihood.

| | | Our Base Model | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Base Sample<br><br>41 countries<br>1995-2019 | | Sample 2<br>(Covid Scenario)<br>41 countries<br>1995-2020 | | Sample 3<br>(Worldwide countries)<br>188 countries<br>1995-2019 | | Base Sample<br>(without GINI)<br>41 countries<br>1995 to 2019 | |
| | Variable | Estimate | P_value | Estimate | P_value | Estimate | P_value | Estimate | P_value |
| Indicator | ServicesGDP | 1.000 | | 1.000 | | 1.000 | | 1.000 | |
| | LaborForceGrowth | -0.004 | 0.593 | -0.032 | 0.143 | -0.034 | 0.000 | -0.010 | 0.309 |
| | Currency | 0.435 | 0.001 | 0.462 | 0.000 | -0.450 | 0.000 | 0.457 | 0.004 |
| | LaborForceParticipation | -0.275 | 0.000 | -0.274 | 0.000 | -0.141 | 0.000 | -0.327 | 0.000 |
| | EnergyConsumptionGrowth | -0.151 | 0.049 | -0.196 | 0.038 | -0.207 | 0.000 | -0.179 | 0.033 |
| Causal | TaxOverGDP | -0.226 | 0.247 | -0.249 | 0.196 | 0.340 | 0.000 | -0.329 | 0.058 |
| | DirectTaxesOverTotalTaxes | 0.179 | 0.007 | 0.175 | 0.007 | -0.108 | 0.000 | 0.074 | 0.145 |
| | SizeOfGovernment | 0.739 | 0.000 | 0.706 | 0.000 | 0.087 | 0.039 | 0.816 | 0.000 |
| | Unemployment | 0.278 | 0.003 | 0.306 | 0.001 | 0.096 | 0.002 | 0.358 | 0.000 |
| | GDPpercapitaPPPgrowth | -0.246 | 0.002 | -0.292 | 0.000 | -0.088 | 0.003 | -0.218 | 0.003 |
| | TradeFreedom | -0.046 | 0.126 | -0.052 | 0.085 | 0.039 | 0.003 | -0.061 | 0.039 |
| | BusinessFreedom | 0.000 | 0.999 | 0.011 | 0.753 | 0.085 | 0.000 | -0.031 | 0.373 |
| | EconomicFreedom | 0.145 | 0.056 | 0.125 | 0.091 | -0.166 | 0.000 | 0.197 | 0.007 |
| | LaborFreedom | -0.149 | 0.000 | -0.154 | 0.000 | 0.000 | 0.980 | -0.157 | 0.000 |
| | LB_GovernmentEffectiveness or WB GovEff | -0.458 | 0.925 | -0.436 | 0.926 | 8.250 | 0.000 | -3.320 | 0.480 |
| | LB_ControlCorruption or WB Control of Corruption | -7.371 | 0.025 | -9.528 | 0.002 | -0.956 | 0.046 | -3.160 | 0.282 |
| | GINI | 0.446 | 0.002 | 0.367 | 0.016 | 0.041 | 0.000 | | |
| | PWT_ctfp_h | 0.199 | 0.000 | 0.186 | 0.000 | 0.077 | 0.000 | 0.160 | 0.000 |
| Statistical tests | RMSEA (p-value <=0.05) | 0.072 | 0.000 | 0.084 | 0.000 | 0.094 | 0.000 | 0.072 | |
| | Chi-square (p-value) | 359,897 | 0.000 | 488,845 | 0.000 | 2,379 | 0.000 | 330,909 | 0.000 |
| | AIC | 77369.48 | | 80513.67 | | 451604.91 | | 75430.68 | |
| | BIC | 78020.56 | | 81169.93 | | 452454.89 | | 76007.78 | |
| | AGFI | 0.962 | | 0.963 | | 0.981 | | 0.951 | |
| | Degrees of freedom | 57 | | 57 | | 57 | | 53 | |
| | Number of missing patterns | 90 | | 110 | | 321 | | 69 | |
| | Number of observations | 1,025 | | 1,066 | | 4,625 | | 1,025 | |

Table 2 – MIMIC Base Model applied to different Samples of countries.

contains highlights on "Estimate" columns for those items that did not meet the expected coefficient signal; and highlights on "P_value" columns for those items that were not found with statistical relevance.

Regarding indicator variables, Currency increases with IE as expected, and Labor Force Participation, Labor Force Growth and Energy Consumption Growth decrease. Labor Force Growth is not statistically significant.

On causal variables, most of the variables are statistically significant, with exception to Tax Over GDP, Business Freedom, and Government Effectiveness. For these variables that are statistically significant, the produced estimator signs are at par with our original expectation except for Economic Freedom and TFP. TFP will be assessed later this section. Economic Freedom in specific is a compounded index from the Heritage Foundation that consolidates other indicators, and we shall consider the possibility of correlation between Heritage's indexes.

Interesting to observe that Sample 3 has further deviations from the expected signs, corroborating to the importance to fine tune MIMIC models to data sets. We observe Currency with a negative signal, and Government Effectiveness with a positive signal, probably results of some customization on variables used for LAC, such as Services GDP and data from LatinoBarometro.

All in all, our Base Model resulted in robust results, with most of the statistically significant estimators producing expected signs. Moreover, choice of variables seems to match well regional data, and not so well Sample 3; be noted that Schneider, Buehn & Montenegro (2010) uses, for instance, different Heritage Foundation indexes, probably those that better match a global model.

Table 3 presents a subset of the results, showing IE estimates for every other 3 years. For each country in the data set we present IE estimates as percentage of GDP from 1998 till 2019, calculated based on the MIMIC coefficients presented on Table 2 with our Base Model applied to the Base Sample of 41 LAC countries from 1995 to 2019.

| Country | 1998 | 2001 | 2004 | 2007 | 2010 | 2013 | 2016 | 2019 | Average | Std.Dev |
|---|---|---|---|---|---|---|---|---|---|---|
| **Argentina** | 24.68 | 25.20 | 21.89 | 22.59 | 22.93 | 23.43 | 23.79 | 23.24 | **23.60** | **4.3%** |
| **Bahamas** | 25.72 | 27.53 | 27.43 | 27.57 | 27.40 | 27.22 | 26.47 | 25.79 | **26.64** | **4.4%** |
| **Barbados** | 29.96 | 30.49 | 30.85 | 25.25 | 27.19 | 24.78 | 23.20 | 25.95 | **27.53** | **10.0%** |
| **Belize** | 41.57 | 44.98 | 46.11 | 45.39 | 43.95 | 43.40 | 43.85 | 45.38 | **44.42** | **2.7%** |
| **Bolivia** | 64.40 | 67.66 | 65.15 | 62.80 | 63.41 | 62.40 | 68.32 | 67.57 | **64.65** | **3.3%** |
| **Brazil** | 40.01 | 38.63 | 37.97 | 39.03 | 38.75 | 40.15 | 41.48 | 41.28 | **39.76** | **2.8%** |
| **Chile** | 19.27 | 19.56 | 18.97 | 19.55 | 19.91 | 20.21 | 20.55 | 20.74 | **19.84** | **2.8%** |
| **Colombia** | 38.72 | 40.87 | 40.36 | 40.83 | 39.28 | 38.71 | 40.05 | 40.94 | **39.61** | **2.5%** |
| **Costa Rica** | 26.06 | 25.24 | 25.58 | 25.04 | 26.75 | 26.83 | 28.12 | 27.35 | **26.35** | **3.5%** |
| **Cuba** | 28.83 | 30.30 | 30.95 | 30.97 | 30.58 | 30.55 | 30.98 | 30.10 | **30.79** | **6.1%** |
| **Dominican Republic** | 32.37 | 34.15 | 34.10 | 33.54 | 34.56 | 34.28 | 33.33 | 32.25 | **33.49** | **2.8%** |
| **Ecuador** | 34.81 | 34.60 | 35.02 | 34.64 | 35.65 | 35.12 | 34.29 | 34.85 | **34.86** | **1.3%** |
| **El Salvador** | 46.12 | 44.62 | 45.02 | 43.37 | 42.97 | 43.37 | 44.06 | 45.17 | **44.36** | **2.3%** |
| **Guatemala** | 51.34 | 50.60 | 50.95 | 52.42 | 53.28 | 53.96 | 54.25 | 54.69 | **52.49** | **2.8%** |
| **Guyana** | 32.44 | 34.20 | 34.41 | 35.67 | 36.24 | 37.07 | 37.19 | 37.32 | **35.47** | **5.7%** |
| **Haiti** | 55.16 | 55.41 | 53.58 | 52.21 | 51.93 | 50.94 | 51.89 | 52.87 | **53.04** | **3.3%** |
| **Honduras** | 48.41 | 49.76 | 51.31 | 53.14 | 52.95 | 52.54 | 51.99 | 51.99 | **51.29** | **3.9%** |
| **Jamaica** | 35.51 | 36.72 | 36.21 | 36.46 | 37.85 | 36.30 | 34.94 | 34.38 | **36.18** | **3.1%** |
| **Mexico** | 29.66 | 32.16 | 32.11 | 31.84 | 32.29 | 32.30 | 32.34 | 32.01 | **31.72** | **2.9%** |
| **Nicaragua** | 45.01 | 43.72 | 43.87 | 45.10 | 43.99 | 42.76 | 43.21 | 43.98 | **44.04** | **2.0%** |
| **Panama** | 63.03 | 67.51 | 64.88 | 67.06 | 66.66 | 64.33 | 63.00 | 62.01 | **64.50** | **2.7%** |
| **Paraguay** | 27.18 | 27.58 | 26.04 | 27.35 | 28.07 | 28.07 | 27.49 | 28.14 | **27.48** | **2.2%** |
| **Peru** | 60.91 | 58.25 | 54.49 | 54.29 | 53.53 | 55.34 | 57.78 | 57.73 | **56.92** | **4.5%** |
| **Puerto Rico** | 28.83 | 29.64 | 29.96 | 31.28 | 30.69 | 31.10 | 30.76 | 30.74 | **30.21** | **3.0%** |
| **Suriname** | 38.41 | 38.58 | 37.30 | 38.06 | 37.89 | 38.60 | 39.94 | 38.46 | **38.16** | **2.8%** |
| **Trinida & Tobago** | 33.58 | 34.68 | 32.47 | 30.87 | 32.11 | 32.83 | 36.60 | 35.30 | **33.25** | **6.1%** |
| **Uruguay** | 49.06 | 52.79 | 50.14 | 47.33 | 48.27 | 48.13 | 49.82 | 50.51 | **49.59** | **3.4%** |
| **Venezuela** | 31.65 | 34.40 | 32.07 | 32.42 | 32.39 | 33.71 | 41.76 | 58.38 | **35.92** | **17.9%** |

Table 3 – MIMIC Base Model IE estimates per country using Base Sample

The average size of IE in the period of 1996-2019 is presented on column “Average” in Table 3. Average values allow us to identify outliers where the informal sector is higher than 50%, as Panama, Bolivia, Guatemala, Haiti, Honduras, and Peru, or higher than 40%, such as Belize, El Salvador, Nicaragua, and Uruguay. On the other hand, countries like Argentina, Bahamas, Barbados, Costa Rica, and Chile present much lower levels of Informal Economy (IE). Estimates are at par with the literature, where we too observe a decline on the levels of IE on the first years of the decade (early 2000s) and a new increase trend from the global financial crisis in 2008 and beyond.

We observe that most countries present an increase in the size of the informal sector from the 2008 financial crisis, at par with results obtained in Matos & Veiga (2019), but with some improvement as the present work also captures a downturn of the informal sector in early 2000s as per Ohnsorge & Yu (2022), as we can see with Argentina, Mexico, Brazil, and Peru in Figure 3. Matos & Veiga (2019) is one of the first works to consider a data set for LAC, and the overall IE growth trend on the past decades is also observed in our work. This long-term growth trend using regional data may be an indicative of the persistence of higher levels of IE due to long term factors such as inequality persistence.

Figure 3 presents the IE estimates obtained with our Base Model, compared to MIMIC and DGE results obtained by the World Bank in their Informal Economy database[2].

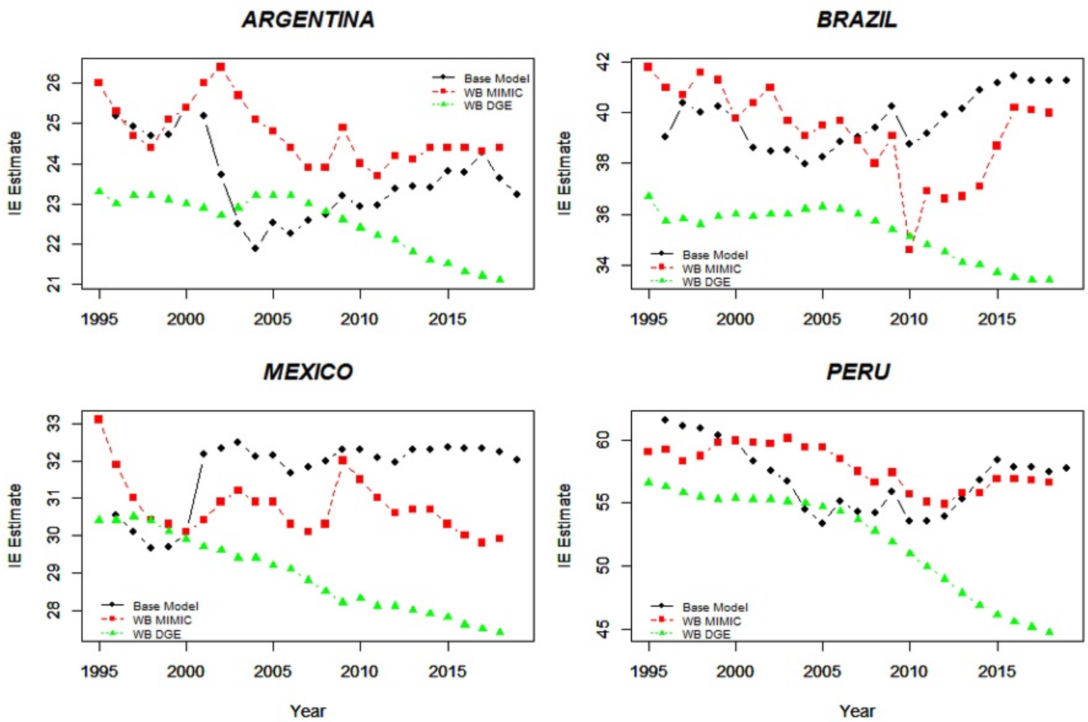


Figure 3 – Comparing MIMIC Base Model IE estimates with World Bank IE database.

In general, results obtained here tend to produce a more pronounced growth trend in IE estimates, as we can see with Brazil, Mexico and Peru, sometimes smoothing peaks and valleys observed on the World Bank MIMIC estimates during the 2008 Financial Crisis. Same behavior is observed in several other countries in the region. The Base Model running on a data set of 41 LAC countries presented here tends to produce IE estimates higher than the World Bank estimates obtained with a data set of 160 countries.

Analyzing Brazil in more detail, we observe a two-percentage-points decline of IE from late-1990s to mid-2000s, which coincides with the realization of several institutional changes made during presidencies from Fernando Henrique Cardoso and early Luis Inácio Lula da Silva. From mid-2000s we observe an increase in the levels of IE to surpass the 40% threshold mid-2010s. Countries like Peru, that had a positive economic cycle later than Brazil, show similar behavior shifted in time, with IE increase more towards late-2010s, when these economies start to show GDP growth reductions. Here we can observe a hint of

2 https://www.worldbank.org/en/research/brief/informal-economy-database

some counter cyclical correlation between formal economic cycles and informal as proposed by Orsi, Raggi & Turino (2014).

In Venezuela we acknowledge an overall degradation of its economy, reflected on the fast increase of IE levels from early 2010s to now, where IE increased from 35% levels to 45%.

We then compared the proposed Base Model with an equivalent model subtracting GINI. Objective here is to try to understand what's the impact of adding GINI to our model. Results show that GINI adds a trend to IE estimates, sometimes upwards, sometimes downwards. Figure 4 presents results for Colombia, Dominican Republic, Argentina, and Mexico, comparing IE estimates for our Base Model (MIMIC 13-1-5) with estimates with the equivalent model excluding GINI as a causal variable (MIMIC 12-1-5).

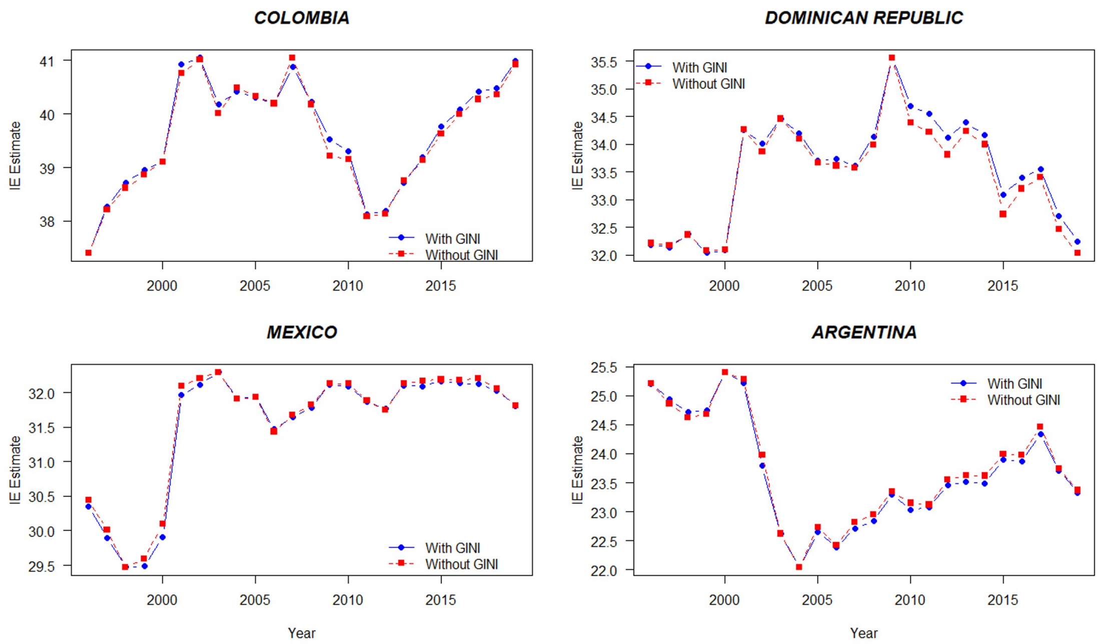


Figure 4 – Comparing Base Model results with and without GINI, with Base Sample of 41 LAC countries, 1995 to 2019.

IE estimates are close, and adding GINI as a causal variable provokes slight changes on the behavior of IE estimates over time. There's no simple rule of thumb that we can infer from the analysis. Here we observe Colombia and Dominican Republic with GINI providing IE estimates in general higher than the model without GINI, while for Mexico and Argentina results with GINI fall short.

Figure 5 presents another comparison of results showing our Base Model with different configurations: fully as presented in Figure 2, excluding only GINI (as compared in Figure 4), excluding only TFP, and excluding both variables. This graph allows us to compare how our model reacts to the addition of GINI and TFP independently, and jointly.

In general, we observe that TFP and GINI exclusions produce results in different directions, what indicates that while GINI increases, usually TFP reduces, and vice-versa, what matches the general intuition. However, GINI seems to have a more consistent trend over time (GUTIÉRREZ-ROMERO, 2021), moving the results from the full Base Model upwards or downwards as observed on Figure 5. TFP, on the other hand, floats upwards and

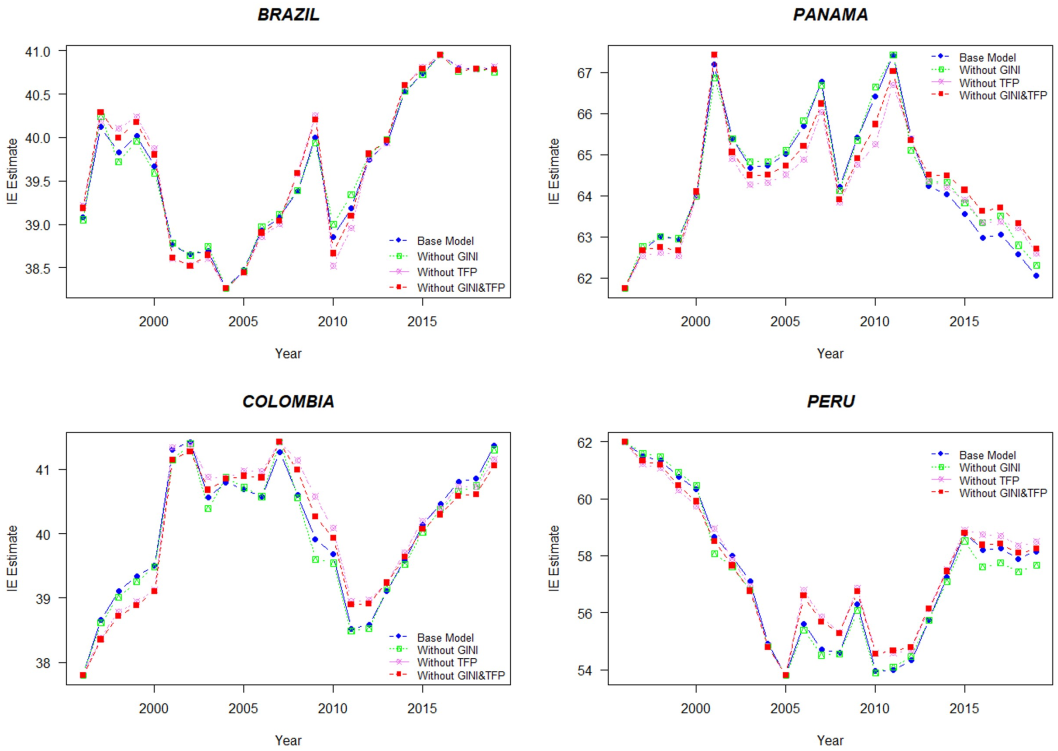


Figure 5 – Comparing Base Model results with and without GINI & TFP, with Base Sample of 41 LAC countries, 1995 to 2019

downwards, representing a more short-term result, what contributes to the hypothesis of TFP working as a proxy to intensive margin. Inequality changes thru GINI are usually observed mid-long term, while TFP changes occur short term.

## 3.1 Covid Impact

We run our Base Model from 1995 to 2019 and compare it to data from 1995 to 2020. Main objective is to assess the initial impact of the Covid-19 pandemic in the informal sector. As expected by the literature negative relation between GDP growth and IE, the huge spike downwards on the economies' GDP growth rate in 2020 has impact resulting on an increase in IE in every country analyzed here. Figure 6 shows the comparison between the two time periods, and it's clear to observe the upward jump of IE estimates in 2020. This will probably be subject of further studies as we get time series covering this period.

From the econometrics standpoint, it's probable that we will have to treat 2020 as an outlier year, as with MIMIC the change on the data set affects the obtained results, as we can see in the graphs. On the other hand, periods such as the 2008 Financial Crisis are not treated as outliers, and they affect the complete estimated IE time series. In the analysis made, some countries show consistence between the two estimates, such as Argentina, Guatemala, and Mexico. Others such as Brazil, Panama and Peru show changes on IE levels that will only be better understood after we have an observation window of some years for the Covid impact.

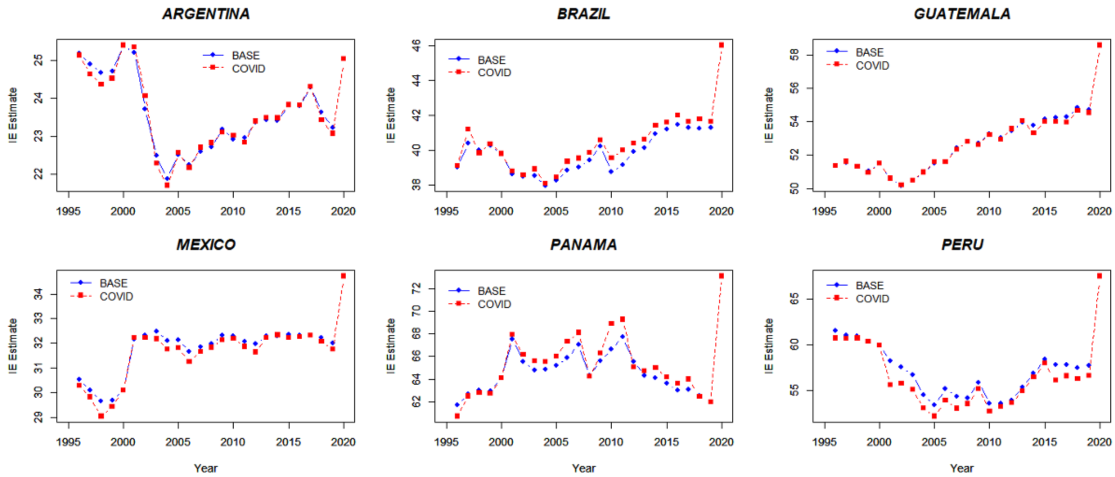


Figure 6 – Covid Impact - BASE scenario (Base Sample of 41 countries, 1995-2019) and COVID scenario (Sample 2 of 41 countries, 1995-2020)

### 3.2 GINI Analysis

GINI showed as an important factor to LAC, where increases on the level of inequality contribute to increases on levels of IE. Alhassan & Haruna (2022) shows equivalent results for Nigeria and proposes that GINI improvements policy making would help on lowering IE.

GINI's behavior on the proposed Base Model as a causal variable is a point of interesting discussion. Despite being object of investigation by some authors such as Chong & Gradstein (2007) and Lesica (2011), recent MIMIC models driven by researchers following Schneider's path have not been using GINI as a proxy of inequality to serve as a causal variable to the informal / shadow economy.

Koufopoulou *et al.* (2021) and Boitano & Deyvi (2019) present GINI as an indicator variable for IE in their studies developed for Greece and Peru, respectively; their conclusion is that an increase in IE produces an increase in GINI, it means, increases on informal economic activity is related to increases on inequality.

Dell'Anno (2021) proposes that higher informality decreases inequality on developing countries, in a bidirectional relationship in which an increase in IE results in further informal labor, with beneficial effects on income distribution, a feedback mechanism that reduces inequality. Berdiev & Saunoris (2019) uses a panel VAR applied to 144 countries over 1960-2009 to present a bidirectional positive relationship between IE and inequality. Gutiérrez-Romero (2021) proposes that past levels of inequality (persistence) are the most important factors explaining the size of IE.

Figure 7 provides a plot of average GINI values and IE estimates from 1999 to 2019 resulting from our Base Model. Countries are aggregated on geographical regions that are indicated with different colors in the graph.

Analyzing the geographical groups separately, we realize that higher absolute GINI values are somehow correlated to higher IE values, in line with the $\gamma$ coefficient positive signal. For Central American countries it gets very apparent, with higher GINI countries such as Belize, Panama, Guatemala, and Honduras corresponding to higher levels of IE. Same happens to South Cone countries, where Brazil and Bolivia demonstrate it while Uruguay

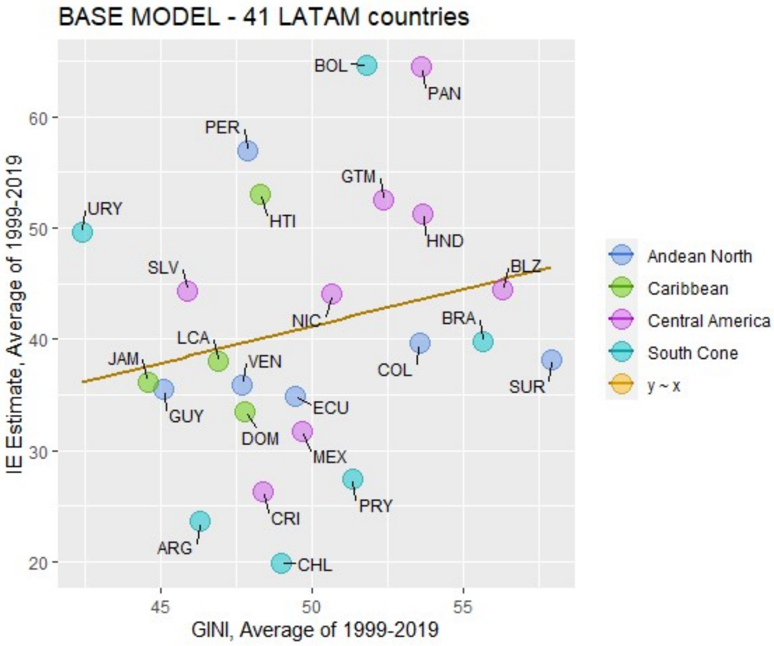


Figure 7 – Average 1999-2019 IE vs GINI for LAC countries

is an outlier with lower relative GINI and high level of IE. The linear regression between GINI and IE results on a positive relation.

In our model, GINI coefficient is relevant and has statistical relevance, and consists of a solid estimator, reflected on lower IE results for Chile and Argentina, and higher for Brazil and Central American countries, for instance. Growth on levels of inequality (higher GINI coefficients) contribute to growth of IE, meaning that with inequality comes challenges to enter the formal market.

On Appendix III we present a discussion on our results correlating GINI and IE, analyzing results or both Base Model and Alternative Model applied to different samples, with all results converging to a positive relation between GINI and IE, corroborating to our findings here. In our analysis we've used GINI as a causal variable to IE, resulting in a positive relation between GINI and Informal Economy.

Additionally, we've exercised our Base Model and our Alternative Model with data from LAC countries, EMDE (Emerging Markets and Developing Economies) and developed countries and the higher and lower IE countries, and we've reached a consistent positive relation between GINI and IE.

## 3.3 TFP Analysis

We proposed TFP as a causal variable to our MIMIC model aiming to use TFP as a driver of GDP growth and economic development, expecting that higher productivity usually tied to more firms' formalization would lead to higher GDP growth and therefore lower IE, it means, a negative relation between TFP and IE. Our results though showed a positive relation between TFP and IE, indicating that higher productivity would lead to higher IE. TFP does not seem to act as a channel for firms' formalization as originally intended.

TFP provides an interesting discussion to our proposed model. As presented by Ulyssea (2020), intensive margin (informal labor) and extensive margin (informal firms) can lead to different directions. In general, research by mid-2010s focused on extensive margin, with causes and consequences tied to firm's decisions whether to formalize or not. Ulyssea (2018) proposes an equilibrium model in which he concludes that firms' informality and informal labor do not move in the same direction in the case of policy changes. Testing the model for the enforced reduction of intensive margin results in an increase of extensive margin, as less productive firms realize higher formalization costs. Additionally, Atesagaoglu, Elgin & Oztunali (2017) shows how TFP and the informal sector are related, comparing results from a one-sector model with results from a two-sector model comprehending formal and informal labor. The two-sector model results in higher levels of TFP, showing how informal labor contributes to higher TFP values.

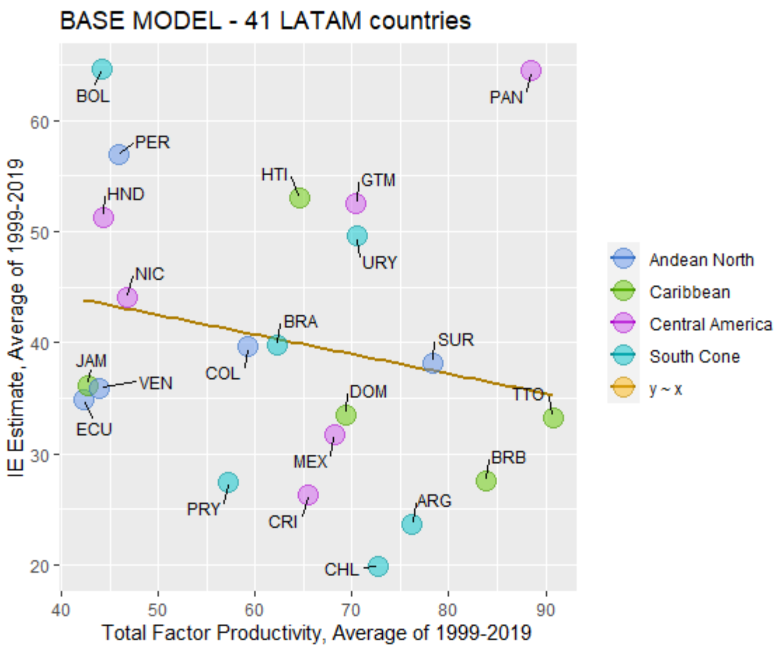


Figure 8 – Average 1999-2019 IE vs TFP for LAC countries

We believe that TFP is a potential channel to the intensive margin or informal labor, although this thesis is not demonstrated here, as it requires different modelling tools to reach such conclusion. Rationale is that TFP results in further formalization of firms and, in parallel, results on increased adoption of informal labor by formal firms. TFP then serves as a bidirectional channel, with negative relation to IE when considering firms formalization, and positive relation to IE when considering informal labor; the positive relation between TFP and IE indicates that labor responds faster to shocks compared to firms' formalization, what's confirmed in the literature.

Figure 8 provides a plot of average TFP values and IE estimates from 1999 to 2019. Although the Base Model results in a positive relation between TFP and IE for each country, the plot in Figure 8 shows a negative relation in a sample of countries. This result reaffirms the general intuition that countries with higher TFP usually are associated with higher GDP and therefore in lower IE. However, for the same country, our study shows a positive relation between TFP and IE, it means if TFP increases in a country, IE increases as well.

The hypothesis that TFP may be a channel for intensive margin needs further study. Probably modelling it with a DSGE/DGE structural model with two sectors, like the one used by Ulyssea (2020), including both extensive and intensive margins would help us confirm our observed results in this study.

### 3.4 Other Results and Robustness Checks

While presenting more detailed results of our Base Model to South Cone, Caribbean, Central America, and Andean geographies on Appendix I, we exercise further robustness checks using different data sets. With the MIMIC methodology we have mainly two main decisions to take:

- **Choice of indicator and causal variables** corresponds to the first main thesis presented in this article, where we propose GINI and TFP as important causal variables to IE. In Appendix II we share an Alternative Model using different variables, with results differing from our Base Model. The Alternative model tends to produce higher estimates for IE, sometimes reaching 3 to 4 percentage points higher than the Base Model, and therefore was deprecated. On the other hand, both GINI and TFP behaves the same in both models.
- **Data Sets** are key to MIMIC estimates. MIMIC methodology requires careful choice of data, and the second main thesis of this article is that estimates for IE in Brazil and countries of LAC are best set with data from the region. Appendix I exercises different data sets and provides an analysis of such exercises, and in general results are similar, but for some specific countries we found a difference of 2 to 3 percentage points on IE estimates. Again, GINI and TFP behavior is consistent as shown hereinbefore.

When applying our Base Model to other geographies in Sample 3 (188 worldwide countries) and other customized samples for EMDE (Emerging Markets and Developing Economies), or Developed Economies, it's interesting to realize that GINI and TFP maintain its nature of statistical relevance and estimation role. For GINI, coefficient values reduce with more heterogeneous countries, such as in Sample 3. Appendix III presents a thorough analysis of GINI vs informality on the different Samples we've built.

## 4 CONCLUSIONS

This article proposes a MIMIC model to estimate the size of the non-observed economy in LAC countries, using a regional specification instead of a global specification. For such we propose causal variables such as GINI representing inequality and TFP representing productivity, both resulting relevant to the region. Results are consistent with the existing literature, and captured economic cycles are in line with expectations. Narrowing the data set down to the region proves to be an effective tool to produce more tailored results, with GINI resulting on a predictor ten-fold more relevant to LAC.

Inequality demonstrated to be an important factor to LAC, where increases of GINI resulted on increases of IE. Transmission channels between inequality and informality need to be considered to avoid policies that aim to reduce informality and result on increasing inequality, as a feedback channel will work to increase back informality. Moreover, GINI resulted as more important to IE when using the regional data set; when running the model with GINI on a global data set, GINI's importance was lower. This result contributes to the hypothesis of leveraging on regional data sets and tailored models to produce improved results.

TFP also showed to be an important factor to LAC, and the positive relation between IE and TFP indicates that intensive margin (labor informality) reacts faster than firms' formalization, consolidating TFP as a candidate channel to informal labor. Long-term policy making should focus on TFP improvements to reach GDP growth and therefore IE reductions, while short-term policies should not disqualify informal labor as a potential channel to improve TFP while also increasing IE.

In LAC, where inequality is already high, and so is informal economy, raising formality seems to be the natural path. However, recent formalization efforts in different countries have demonstrated that such task has no easy solution; in several cases formality rate increase has been acquired at the expense of no incremental tax collection and without improving firm's productivity after formalization (DELL'ANNO, 2021). Usage of GINI as a causal variable in our proposed base model reinforces the general hypothesis that reducing inequality is a strong mechanism to reduce informality on the economy.

Informality in our region is also tied to a burdensome tax system and rigid labor markets, as indicated by our model, confirming literature on the topic. On the other hand, informality, especially on the labor market, reduces labor unionization, overall tax burden and social contribution by firms, eventually resulting in improved overall system efficiency, on the contrary of common sense. In regions such as LAC, where informal and formal activities are intertwined, it's worth further exploring the relations between inequality and informality.

Inequality results obtained here have key implications for policy recommendations in LAC countries and the required methodology to estimate informality. All in all, policymakers should focus on reducing inequality in the mid-long term, as it contributes to reducing IE. However, in the short-term IE is an important economic buffer to shocks, especially because of informal labor in LAC, and hence policy making to safe net informal labor may be a path to be pursued.

# REFERENCES

ALHASSAN, U.; HARUNA, E. U. Demystifying rising income inequality influence on shadow economy: Empirical evidence from Nigeria. *Review of Economic Analysis*, v. 14, n. 2, p. 293–318, 2022. 14

ATESAGAOGLU, O. E.; ELGIN, C.; OZTUNALI, O. TFP growth in Turkey revisited: The effect of informal sector. *Central Bank Review*, Elsevier, v. 17, n. 1, p. 11–17, 2017. 16

BERDIEV, A. N.; SAUNORIS, J. W. On the relationship between income inequality and the shadow economy. *Eastern Economic Journal*, Springer, v. 45, n. 2, p. 224–249, 2019. 14

BOITANO, G.; DEYVI, F. The informal economy and its impact on tax revenues and economic growth. the case of Peru, Latin America and OECD countries (1995 – 2016).

*360 Revista de Ciencias de la Gestión*, v. 4, n. 4, p. 128–157, 2019. 14

BREUSCH, T. *Estimating the Underground Economy using MIMIC Models*. [S.l.], 2005. Disponível em: <https://EconPapers.repec.org/RePEc:wpa:wuwpem:0507003>. 5, 7

CHANG, C. *et al.* Comparing multiple statistical software for multiple-indicator, multiple-cause modeling: an application of gender disparity in adult cognitive functioning using MIDUS II dataset. *BMC Medical Research Methodology*, Springer, v. 20, n. 1, p. 1–14, 2020. 8

CHEN, M. A. *The informal economy: Definitions, theories and policies*. [S.l.], 2012. 2

CHONG, A.; GRADSTEIN, M. Inequality and informality. *Journal of public Economics*, Elsevier, v. 91, n. 1-2, p. 159–179, 2007. 14

DELL'ANNO, R. Inequality and informality in transition and emerging countries. *IZA World of Labor*, 2021. 14, 18

DELL'ANNO, R. Theories and definitions of the informal economy: A survey. *Journal of Economic Surveys*, Wiley Online Library, v. 36, n. 5, p. 1610–1643, 2022. 2

DELL'ANNO, R.; SCHNEIDER, F. Estimating the underground economy by using MIMIC models: A response to T. Breusch´ s critique. 2006. 5

DELL'ANNO, R.; SCHNEIDER, F. A complex approach to estimate shadow economy: the structural equation modelling. In: *Coping with the Complexity of Economics*. [S.l.]: Springer, 2009. p. 111–130. 5

GUTIÉRREZ-ROMERO, R. Inequality, persistence of the informal economy, and club convergence. *World Development*, Elsevier, v. 139, p. 105211, 2021. 3, 12, 14

ILO. *Conceptual framework for statistics on informal economy*. [S.l.], 2021. Disponível em: <https://www.ilo.org/ilostat-files/Documents/Informality_conceptual_framework_draft.pdf>. 3

JÖRESKOG, K. G.; GOLDBERGER, A. S. Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American statistical Association*, Taylor & Francis, v. 70, n. 351a, p. 631–639, 1975. 3

KOUFOPOULOU, P. *et al.* Estimating shadow economy size in Greece 2000-2018: A flexible MIMIC approach. *SPOUDAI-Journal of Economics and Business*, v. 71, n. 3-4, p. 23–47, 2021. 14

LESICA, G. T. Estimating the size of the informal sector in Latin America. 2011. 14

MATOS, P.; VEIGA, F. On the Shadow Economy in Latin America. *International Journal of Applied Economics*, p. 7–21, 2019. 2, 3, 11, 23, 27

MEDINA, L.; SCHNEIDER, F. *Shedding Light on the Shadow Economy: A Global Database and the Interaction with the Official One*. [S.l.], 2019. Disponível em: <https://EconPapers.repec.org/RePEc:ces:ceswps:_7981>. 3

NOGUEIRA, M. TD 2237 - A Construção Social da Informalidade e da Semiformalidade no Brasil: uma proposta para o debate. *Texto Para Discussão - Ipea*, v. 2233, 10 2016. 2

OECD. *Measuring the non-observed economy - A handbook organization for economic cooperation and development.* [S.l.], 2002. 2

OHNSORGE, F.; YU, S. *The Long Shadow of Informality: Challenges and Policies.* Washington, DC: World Bank Publications, 2022. 5, 11

ORSI, R.; RAGGI, D.; TURINO, F. Size, trend, and policy implications of the underground economy. *Review of Economic Dynamics*, Elsevier, v. 17, n. 3, p. 417–436, 2014. 12

PETERS, A. Estimating the size of the informal economy in Caribbean States. IADB: Inter-American Development Bank, 2017. 22

RAUCH, J. E. Modelling the informal sector formally. *Journal of Development Economics*, Elsevier, v. 35, n. 1, p. 33–47, 1991. 3

SCHNEIDER, F. Shadow Economies and Corruption All Over the World: New Estimates for 145 Countries. *Economics - The Open-Access, Open-Assessment E-Journal (2007-2020)*, v. 1, p. 1–66, 2007. Disponível em: <https://EconPapers.repec.org/RePEc:zbw:ifweej:5744>. 8

SCHNEIDER, F.; BUEHN, A. Estimating the size of the shadow economy: Methods, problems and open questions. CESifo Working Paper Series, 2013. 5

SCHNEIDER, F.; BUEHN, A.; MONTENEGRO, C. E. New estimates for the shadow economies all over the world. *International Economic Journal*, Taylor & Francis, v. 24, n. 4, p. 443–461, 2010. 2, 6, 8, 10

SCHNEIDER, F.; ENSTE, D. H. Shadow economies: Size, causes, and consequences. *Journal of economic literature*, v. 38, n. 1, p. 77–114, 2000. 3

SOTO, H. d. *The other path : the economic answer to terrorism.* New York: Basic Books, 1989. ISBN 0465016103. 3

TANZI, V. The Shadow Economy, Its Causes and Its Consequences. 01 1999. 3

TREBICKA, B. MIMIC Model: A Tool to Estimate the Shadow Economy. *Academic Journal of Interdisciplinary Studies*, Rome, Italy, v. 3, n. 6, 2014. 5

ULYSSEA, G. Firms, informality, and development: Theory and evidence from Brazil. *American Economic Review*, v. 108, n. 8, p. 2015–47, 2018. 16

ULYSSEA, G. Informality: Causes and consequences for development. *Annual Review of Economics*, Annual Reviews, p. 525–46, 2020. 16, 17

# A APPENDIX I - BASE MODEL APPLIED TO SAMPLES

Table 4 provides the complete results from our proposed Base Model for LAC. It contains the yearly data resulting from our estimates extending the subset results presented in Table 3. Values are presented as percentage points of GDP.

| Country | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | Average | SD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Argentina | 25.19 | 24.91 | 24.68 | 24.72 | 25.40 | 25.20 | 23.72 | 22.49 | 21.89 | 22.53 | 22.25 | 22.59 | 22.72 | 23.20 | 22.93 | 22.97 | 23.37 | 23.43 | 23.41 | 23.82 | 23.79 | 24.29 | 23.64 | 23.24 | 23.60 | 4.3% |
| Bahamas | 23.91 | 24.06 | 25.72 | 25.27 | 26.20 | 27.53 | 27.12 | 28.00 | 27.43 | 27.78 | 26.76 | 27.57 | 27.78 | 28.57 | 27.40 | 27.53 | 25.91 | 27.22 | 26.90 | 26.74 | 26.47 | 25.80 | 25.99 | 25.79 | 26.64 | 4.4% |
| Barbados | 30.31 | 31.16 | 29.96 | 30.04 | 30.00 | 30.49 | 30.87 | 29.97 | 30.85 | 30.93 | 26.80 | 25.25 | 23.64 | 26.01 | 27.19 | 24.61 | 25.96 | 24.78 | 28.18 | 24.71 | 23.20 | 25.19 | 24.64 | 25.95 | 27.53 | 10.0% |
| Belize | 43.09 | 43.59 | 41.57 | 43.62 | 43.80 | 44.98 | 46.20 | 46.18 | 46.11 | 45.74 | 44.54 | 45.39 | 44.90 | 44.84 | 43.95 | 44.42 | 43.76 | 43.40 | 43.25 | 43.01 | 43.85 | 44.24 | 46.26 | 45.38 | 44.42 | 2.7% |
| Bolivia | 63.27 | 64.33 | 64.40 | 66.68 | 67.10 | 67.66 | 66.50 | 64.78 | 65.15 | 64.73 | 62.13 | 62.80 | 61.95 | 63.00 | 63.41 | 61.67 | 62.52 | 62.40 | 62.09 | 66.52 | 68.32 | 67.14 | 65.40 | 67.57 | 64.65 | 3.3% |
| Brazil | 39.04 | 40.40 | 40.01 | 40.26 | 39.80 | 38.63 | 38.48 | 38.53 | 37.97 | 38.25 | 38.84 | 39.03 | 39.43 | 40.23 | 38.75 | 39.17 | 39.90 | 40.15 | 40.92 | 41.20 | 41.48 | 41.29 | 41.26 | 41.28 | 39.76 | 2.8% |
| Chile | 18.81 | 19.16 | 19.27 | 19.71 | 19.80 | 19.56 | 19.51 | 19.36 | 18.97 | 19.90 | 19.22 | 19.55 | 20.01 | 20.31 | 19.91 | 19.53 | 20.09 | 20.21 | 20.49 | 20.62 | 20.55 | 20.38 | 20.53 | 20.74 | 19.84 | 2.8% |
| Colombia | 37.43 | 38.28 | 38.72 | 38.94 | 39.10 | 40.87 | 40.99 | 40.14 | 40.36 | 40.27 | 40.15 | 40.83 | 40.18 | 39.51 | 39.28 | 38.14 | 38.19 | 38.71 | 39.18 | 39.73 | 40.05 | 40.38 | 40.43 | 40.94 | 39.61 | 2.5% |
| Costa Rica | 26.31 | 27.10 | 26.06 | 26.05 | 26.20 | 25.24 | 25.53 | 25.42 | 25.58 | 25.00 | 25.28 | 25.04 | 25.58 | 26.10 | 26.75 | 26.95 | 26.19 | 26.83 | 26.86 | 27.35 | 28.12 | 28.00 | 27.60 | 27.35 | 26.35 | 3.5% |
| Cuba | 39.11 | 29.14 | 28.83 | 29.76 | 30.00 | 30.30 | 30.55 | 30.82 | 30.95 | 31.28 | 31.29 | 30.97 | 30.62 | 30.57 | 30.58 | 31.00 | 30.26 | 30.55 | 30.68 | 30.65 | 30.98 | 30.02 | 29.90 | 30.10 | 30.79 | 6.1% |
| Dominican Republic | 32.19 | 32.16 | 32.37 | 32.05 | 32.10 | 34.15 | 33.92 | 34.35 | 34.10 | 33.63 | 33.66 | 33.54 | 34.04 | 35.38 | 34.56 | 34.43 | 34.02 | 34.28 | 34.07 | 33.05 | 33.33 | 33.48 | 32.69 | 32.25 | 33.49 | 2.8% |
| Ecuador | 34.51 | 34.73 | 34.81 | 34.68 | 34.40 | 34.60 | 35.28 | 35.92 | 35.02 | 35.18 | 34.41 | 34.64 | 34.18 | 35.66 | 35.65 | 35.34 | 34.91 | 35.12 | 34.86 | 34.72 | 34.29 | 34.15 | 34.64 | 34.85 | 34.86 | 1.3% |
| El Salvador | 44.30 | 45.62 | 46.12 | 45.79 | 46.30 | 44.62 | 45.03 | 44.24 | 45.02 | 44.37 | 43.90 | 43.37 | 43.50 | 44.18 | 42.97 | 42.70 | 42.87 | 43.37 | 43.48 | 43.96 | 44.06 | 44.67 | 44.93 | 45.17 | 44.36 | 2.3% |
| Guatemala | 51.36 | 51.54 | 51.34 | 51.04 | 51.50 | 50.60 | 50.19 | 50.48 | 50.95 | 51.53 | 51.59 | 52.42 | 52.83 | 52.69 | 53.28 | 53.03 | 53.48 | 53.96 | 53.77 | 54.17 | 54.25 | 54.27 | 54.83 | 54.69 | 52.49 | 2.8% |
| Guyana | 31.74 | 31.48 | 32.44 | 32.80 | 33.40 | 34.20 | 34.12 | 34.39 | 34.41 | 35.36 | 36.25 | 35.67 | 36.26 | 36.25 | 36.24 | 36.95 | 36.70 | 37.07 | 37.70 | 37.46 | 37.19 | 37.77 | 38.14 | 37.32 | 35.47 | 5.7% |
| Haiti | 53.71 | 54.72 | 55.16 | 55.90 | 55.40 | 55.41 | 55.84 | 55.06 | 53.58 | 53.82 | 52.85 | 52.21 | 51.94 | 51.85 | 51.93 | 50.85 | 50.66 | 50.94 | 51.22 | 51.20 | 51.89 | 52.17 | 51.89 | 52.87 | 53.04 | 3.3% |
| Honduras | 47.76 | 47.54 | 48.41 | 48.43 | 49.60 | 49.76 | 50.71 | 51.07 | 51.31 | 51.65 | 53.03 | 53.14 | 54.12 | 54.31 | 52.95 | 53.41 | 53.82 | 52.54 | 51.88 | 51.39 | 51.99 | 50.47 | 49.70 | 51.99 | 51.29 | 3.9% |
| Jamaica | 33.86 | 35.08 | 35.51 | 35.92 | 36.40 | 36.72 | 37.54 | 37.35 | 36.21 | 36.44 | 36.23 | 36.46 | 36.89 | 37.87 | 37.85 | 37.36 | 37.46 | 36.30 | 36.09 | 35.63 | 34.94 | 34.92 | 34.94 | 34.38 | 36.18 | 3.1% |
| Mexico | 30.54 | 30.09 | 29.66 | 29.67 | 30.10 | 32.16 | 32.32 | 32.49 | 32.11 | 32.15 | 31.67 | 31.84 | 31.98 | 32.31 | 32.29 | 32.07 | 31.97 | 32.30 | 32.29 | 32.37 | 32.34 | 32.33 | 32.23 | 32.01 | 31.72 | 2.9% |
| Nicaragua | 45.50 | 44.64 | 45.01 | 44.77 | 45.20 | 43.72 | 43.79 | 44.24 | 43.87 | 43.83 | 45.12 | 45.10 | 44.83 | 44.97 | 43.99 | 42.85 | 42.99 | 42.76 | 42.73 | 42.56 | 43.21 | 43.45 | 43.81 | 43.98 | 44.04 | 2.0% |
| Panama | 61.69 | 62.67 | 63.03 | 62.96 | 64.10 | 67.51 | 65.58 | 64.82 | 64.88 | 65.18 | 65.89 | 67.06 | 64.31 | 65.59 | 66.66 | 67.74 | 65.57 | 64.33 | 64.13 | 63.61 | 63.00 | 63.07 | 62.57 | 62.01 | 64.50 | 2.7% |
| Paraguay | 26.45 | 26.91 | 27.18 | 27.42 | 27.40 | 27.58 | 26.82 | 26.91 | 26.04 | 26.87 | 27.94 | 27.35 | 27.21 | 27.67 | 28.07 | 27.64 | 27.61 | 28.07 | 28.51 | 28.44 | 27.49 | 28.03 | 27.77 | 28.14 | 27.48 | 2.2% |
| Peru | 61.56 | 61.05 | 60.91 | 60.35 | 59.90 | 58.25 | 57.57 | 56.68 | 54.49 | 53.39 | 55.18 | 54.29 | 54.18 | 55.86 | 53.53 | 53.56 | 53.91 | 55.34 | 56.84 | 58.36 | 57.78 | 57.82 | 57.46 | 57.73 | 56.92 | 4.5% |
| Puerto Rico | 29.04 | 28.78 | 28.83 | 28.29 | 28.40 | 29.64 | 29.83 | 30.01 | 29.96 | 30.78 | 31.08 | 31.28 | 31.23 | 30.63 | 30.69 | 30.59 | 30.57 | 31.10 | 30.84 | 30.49 | 30.76 | 30.73 | 30.72 | 30.74 | 30.21 | 3.0% |
| Suriname | 35.63 | 36.62 | 38.41 | 39.77 | 39.80 | 38.58 | 38.14 | 38.52 | 37.30 | 37.52 | 38.13 | 38.06 | 37.22 | 38.30 | 37.89 | 38.32 | 37.45 | 38.60 | 39.17 | 39.53 | 39.94 | 36.81 | 37.66 | 38.46 | 38.16 | 2.8% |
| Trinidad & Tobago | 32.92 | 32.81 | 33.58 | 34.53 | 34.40 | 34.68 | 35.23 | 32.79 | 32.47 | 30.43 | 29.79 | 30.87 | 29.40 | 32.20 | 32.11 | 31.79 | 32.38 | 32.83 | 33.23 | 35.74 | 36.60 | 36.54 | 35.46 | 35.30 | 33.25 | 6.1% |
| Uruguay | 51.71 | 51.02 | 49.06 | 48.47 | 51.10 | 52.79 | 53.21 | 51.14 | 50.14 | 48.95 | 48.94 | 47.33 | 47.71 | 47.92 | 48.27 | 47.71 | 48.29 | 48.13 | 47.83 | 48.46 | 49.82 | 51.15 | 50.48 | 50.51 | 49.59 | 3.4% |
| Venezuela | 33.92 | 32.27 | 31.65 | 33.06 | 33.60 | 34.40 | 34.83 | 34.19 | 32.07 | 31.64 | 31.41 | 32.42 | 32.12 | 34.12 | 32.39 | 32.13 | 33.72 | 33.71 | 36.13 | 39.35 | 41.76 | 44.35 | 48.53 | 58.38 | 35.92 | 17.9% |

Table 4 – MIMIC Base Model yearly IE estimates per country, from 1996 to 2019. Complement of Table 2

In Table 5 we show the statistics for causal and indicator variables used in our model, as well as the estimators obtained with our Base Model, with and without GINI & TFP.

| | | | | | | | BASE MODEL 41 countries 1995 to 2019 | | BASE MODEL 41 countries 1995 to 2019 w/o GINI & TFP | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Variable | # Observations | Average | Variance | MIN | MAX | Estimate | P_value | Estimate | P_value |
| Indicator | ServicesGDP | 878 | 59.12 | 11.20 | 21.88 | 90.82 | 1 | | 1 | |
| | LaborForceGrowth | 775 | 1.77 | 1.94 | 8.88 | 10.65 | -0.004 | 0.593 | -0.003 | 0.857 |
| | Currency | 515 | 46.65 | 24.30 | 0.05 | 100.00 | 0.435 | 0.001 | 0.838 | 0.000 |
| | LaborForceParticipation | 750 | 61.89 | 7.07 | 40.36 | 79.75 | -0.275 | 0.000 | -0.518 | 0.000 |
| | EnergyConsumptionGrowth | 1024 | 3.76 | 17.54 | 47.59 | 374.22 | -0.151 | 0.049 | -0.269 | 0.020 |
| Causal | TaxOverGDP | 476 | 15.89 | 4.91 | 7.71 | 30.26 | -0.226 | 0.247 | -0.460 | 0.005 |
| | DirectTaxesOverTotalTaxes | 652 | 28.58 | 10.84 | 4.05 | 76.12 | 0.179 | 0.007 | 0.163 | 0.001 |
| | SizeOfGovernment | 1000 | 15.00 | 7.42 | 3.95 | 57.47 | 0.739 | 0.000 | 0.706 | 0.000 |
| | Unemployment | 775 | 8.62 | 4.73 | 1.58 | 22.21 | 0.278 | 0.003 | 0.288 | 0.003 |
| | GDPpercapitaPPPgrowth | 886 | 3.74 | 4.16 | 12.67 | 23.59 | -0.246 | 0.002 | -0.140 | 0.050 |
| | TradeFreedom | 685 | 66.32 | 17.37 | 0.00 | 90.00 | -0.046 | 0.126 | -0.074 | 0.008 |
| | BusinessFreedom | 689 | 62.19 | 17.46 | 0.00 | 99.90 | 0.000 | 0.999 | -0.045 | 0.147 |
| | EconomicFreedom | 685 | 59.26 | 12.30 | 1.00 | 84.20 | 0.145 | 0.056 | 0.213 | 0.003 |
| | LaborFreedom | 459 | 60.30 | 16.86 | 0.00 | 100.00 | -0.149 | 0.000 | -0.199 | 0.000 |
| | LB_GovernmentEffectiveness | 339 | 0.34 | 0.15 | 0.05 | 0.84 | -0.458 | 0.925 | 4.701 | 0.201 |
| | LB_ControlCorruption | 322 | 0.71 | 0.21 | 0.17 | 0.93 | -7.371 | 0.025 | -3.630 | 0.151 |
| | GINI | 337 | 49.85 | 5.02 | 38.00 | 61.60 | 0.446 | 0.002 | | |
| | PWT_ctfp_h | 500 | 62.16 | 19.94 | 5.44 | 166.70 | 0.199 | 0.000 | | |

Table 5 – MIMIC Base Model with indicator and causal variables statistics & results

We present in Figures 9 and 10 the results of our proposed Base Model to different countries, from 1995 to 2019. Figure 9 shows the results for countries in Central America and the Caribbean.

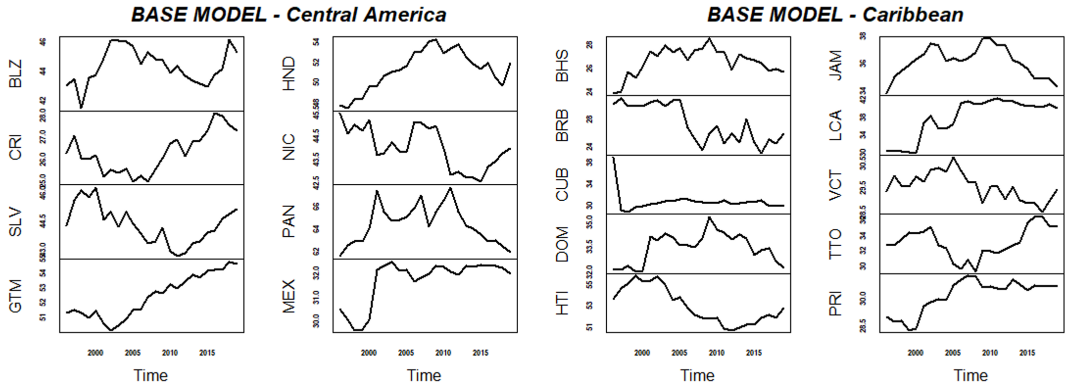


Figure 9 – MIMIC Base Model estimates for IE – Central America and Caribbean countries - Base Sample of 41 LAC countries 1995-2019

Central American countries have traditionally faced high levels of informality, with the exception to Costa Rica (CRI). Behavior is different country-by-country, but in general we see an upwards trend from mid 2000s. Caribbean countries such as Bahamas (BHS) and Barbados (BRB) show relative lower levels of informality, while countries like Haiti (HTI) shows levels similar to what we observe in Central America. For Caribbean countries, results are at par with Peters (2017).

Figure 10 shows the estimates for the South Cone and Andean & North countries. Argentina (ARG), Bolivia (BOL), Brazil (BRA), Colombia (COL) and Peru (PER) show a similar trend, with IE decline from early-2000s to mid-2000s and increase of IE levels happening towards the financial crisis. Such behavior is not so clear with Matos & Veiga (2019), which also leverages on a regional data set.

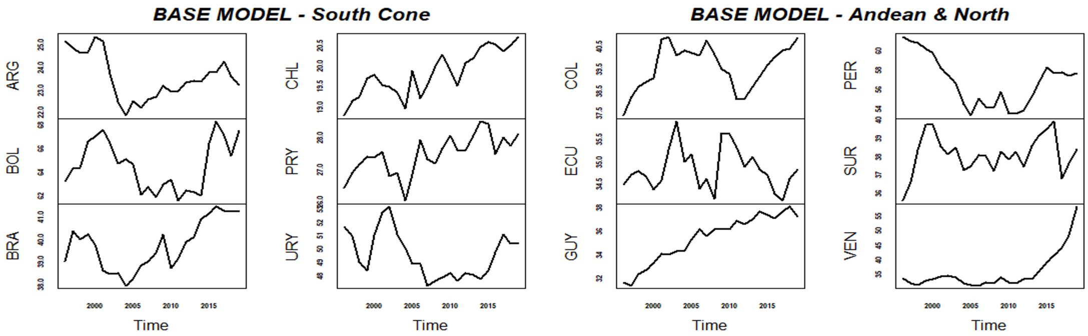


Figure 10 – MIMIC Base Model estimates for IE – South Cone and Andean & North countries - Base Sample of 41 LAC countries 1995-2019

Figure 11 presents the results shown in the figures above in a different format, for some countries that were not presented on Figure 3. On these graphs we can compare the predictions obtained with our proposed Base Model and the World Bank estimates with MIMIC and DGE. Like what we've seen on Figure 3, our Base Model estimates produce results at par with the literature, and deviations cannot be explained with a simple rule of thumb. Venezuela shows up as a significant spike in the recent decade with our model, which seems to better reflect country conditions.

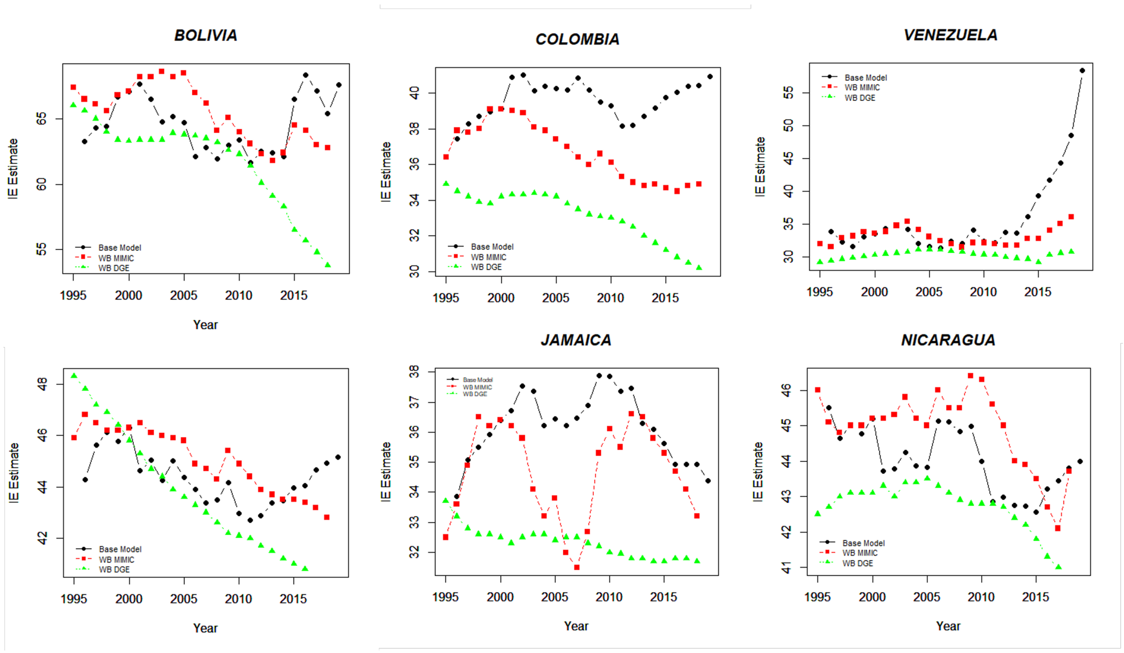


Figure 11 – Robustness Check for MIMIC Base Model IE estimates compared to World Bank Database

Figure 12 shows IE estimates with our Base Model compared to Sample 3, applied to some main LAC countries (Argentina, Brazil, Colombia, Peru, Venezuela, Mexico, Guatemala, Panama, and Puerto Rico), for the period of 1995 to 2019. For some countries the behavior of the models is similar, like Argentina, but in general the two models produce very different results.

Sample 3 applies the Base Model to an extensive set of 188 countries. As LatinoBarometro data is not available for this broader data set, we exchanged LB-GovernmentEffectiveness by World Bank's WGI Government Effectiveness variable (GovEff). LB-ControlCorruption was replaced by World Bank's WGI Control of Corruption variable. Results present higher statistical significance as the sample of data is much bigger. The behavior of both GINI and TFP is similar to the Base Model.

This exercise corroborates the importance of data set selection on MIMIC estimates, as data changes the prediction of the same MIMIC model significantly. Landing our work in this article with a sample of 41 LAC countries reinforce the importance of the data set selection while using MIMIC.

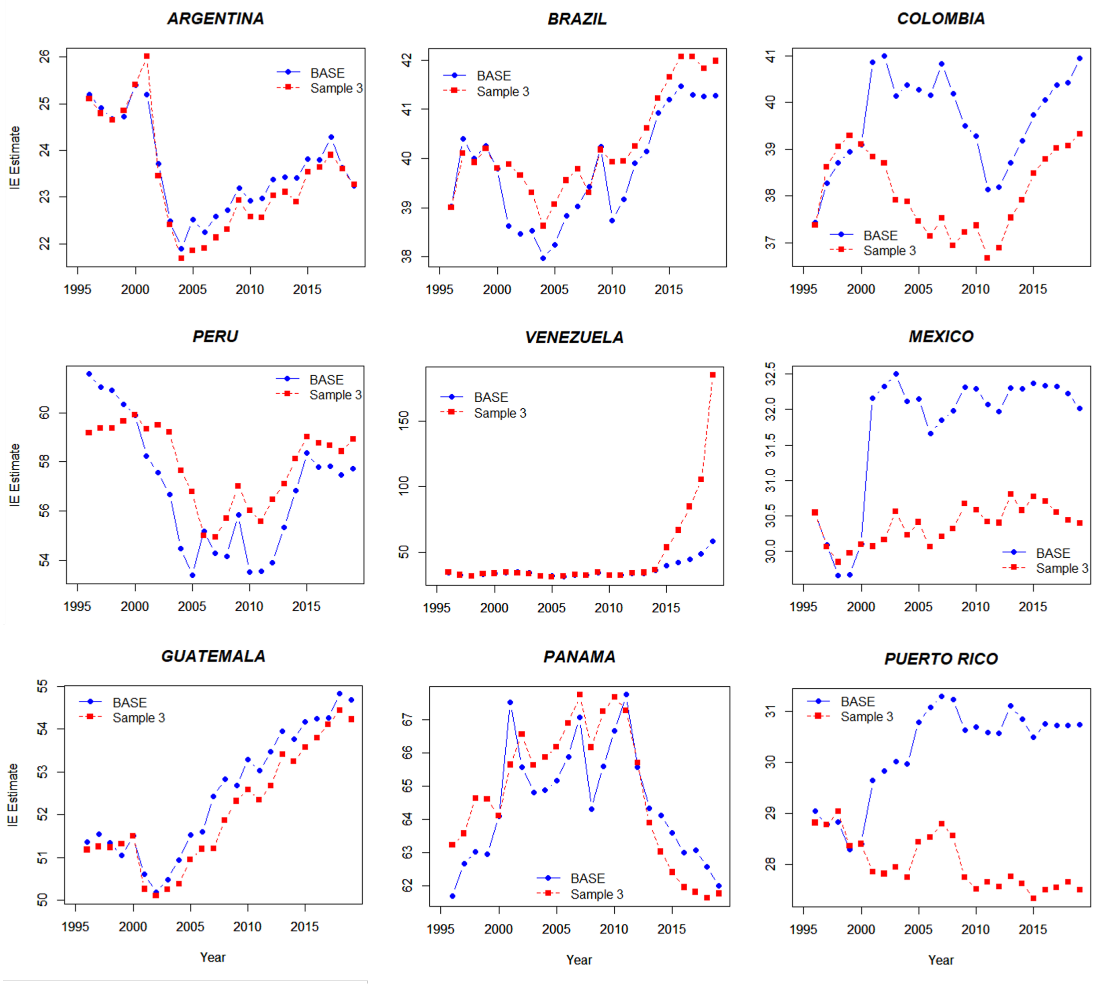


Figure 12 – Robustness Check for Base Model applied to Base Sample (41 LAC countries) and applied to Sample 3 (188 countries)

For Mexico, for instance, usage of the LAC data set results in an informal sector almost two percentile points of GDP higher. Puerto Rico shows similar behavior with a difference of three percentile points.

There's no systematic rule when we compare the models. In some countries such as Brazil and Peru, the model with LAC data set produces lower IE estimates while in countries such as Mexico, Puerto Rico, Colombia, and Guatemala it produces higher estimates.

Venezuela presents an interesting situation. With Sample 3, IE estimates reach more than 150 percentile points, meaning the informal sector produces more than the formal one.

# B APPENDIX II - ALTERNATIVE MODEL

In the search of a model that could produce results for inequality and TFP, as presented in our Base Model, we have exercised several intermediate models that presented different behaviors. One interesting model identified is presented in Table 6, named Alternative Base Model.

| | Variable | # Observations | Average | Variance | MIN | MAX | Estimate | P_value |
|---|---|---|---|---|---|---|---|---|
| Indicator | ServicesGDP | 878 | 59.12 | 11.20 | 21.88 | 90.82 | 1 | |
| Indicator | LaborForceGrowth | 775 | 1.77 | 1.94 | 8.88 | 10.65 | -0.083 | 0.000 |
| Indicator | Broad Money | 800 | 54.48 | 24.38 | 13.15 | 150.71 | 3.468 | 0.000 |
| Indicator | LaborForceParticipation | 750 | 61.89 | 7.07 | 40.36 | 79.75 | -0.162 | 0.001 |
| Causal | TaxOverGDP | 476 | 15.89 | 4.91 | 7.71 | 30.26 | 0.643 | 0.000 |
| Causal | SizeOfGovernment | 1000 | 15.00 | 7.42 | 3.95 | 57.47 | 0.293 | 0.000 |
| Causal | Unemployment | 775 | 8.62 | 4.73 | 1.58 | 22.21 | 0.290 | 0.000 |
| Causal | GDPpercapitaPPPgrowth | 886 | 3.74 | 4.16 | 12.67 | 23.59 | 0.000 | 0.993 |
| Causal | TradeOpen | 1000 | 85.60 | 37.71 | 15.61 | 200.73 | -0.033 | 0.002 |
| Causal | Credit | 728 | 42.07 | 20.82 | 6.59 | 123.82 | 0.249 | 0.000 |
| Causal | Interest | 747 | 9.40 | 11.55 | -62.75 | 93.92 | -0.048 | 0.056 |
| Causal | EIU_RegQual | 359 | 0.57 | 0.18 | 0.05 | 0.90 | -7.514 | 0.002 |
| Causal | EIU_GE_Bureauc | 359 | 0.35 | 0.23 | 0.00 | 0.75 | -5.710 | 0.025 |
| Causal | LB_GovernmentEffectiveness | 339 | 0.34 | 0.15 | 0.05 | 0.84 | -0.668 | 0.777 |
| Causal | LB_ControlCorruption | 322 | 0.71 | 0.21 | 0.17 | 0.93 | -4.547 | 0.011 |
| Causal | GINI | 337 | 49.85 | 5.02 | 38.00 | 61.60 | 0.125 | 0.189 |
| Causal | PWT_ctfp_h | 500 | 62.16 | 19.94 | 5.44 | 166.70 | 0.093 | 0.000 |
| Causal | hc | 600 | 2.47 | 0.41 | 1.43 | 3.61 | -3.435 | 0.000 |
| Statistical tests | RMSEA (p-value <=0.05) | | | | | | 0.112 | |
| Statistical tests | Chi-square (p-value) | | | | | | 609,607 | |
| Statistical tests | AIC | | | | | | 67723.63 | |
| Statistical tests | BIC | | | | | | 68438.84 | |
| Statistical tests | AGFI | | | | | | 0.910 | |
| Statistical tests | Degrees of freedom | | | | | | 44 | |
| Statistical tests | Number of missing patterns | | | | | | 103 | |
| Statistical tests | Number of observations | | | | | | 1,025 | |

Table 6 – Alternative Base Model results

This model contains the following characteristics:

- Reduced indicator variables from five to four, resulting in a MIMIC 14-1-4 instead of a MIMIC 13-1-5 corresponding to the Base Model.
- We do not use the time series data that were manipulated: Currency and DirectTaxesOverTotalTaxes. The former is replaced by Broad Money, and the latter is not used as a causal variable.
- We replace all Governance indicators per a different set of indicators:
  - We use Trade Openness in lieu of Trade Freedom
  - EIU's Regulatory Quality replaces Labor Freedom
  - EIU's Government Efficiency Bureaucracy replaces Business Freedom

  – Credit and Interest replaces Economic Freedom.

- We include PWT's Human Capital (hc) index, which combines the average years of schooling from Barro & Lee and an education rate of return based on Mincer equation.

The Alternative Base Model was discarded because it did not reach GINI's statistical relevance, nevertheless it has the merit to maintain importance to TFP, additionally adding the Human Capital Index with a negative sign, meaning that higher levels of HC result in decrease of IE, as expected.

Figure 13 shows the results of the Alternative Model compared to our Base Model. The Alternative Model differs significantly from the Base Model, presenting in several countries an upward trend in the analyzed period. As reviewed in the Results session, the Base Model provides better estimates if we analyze the economics history of each country. In general, the Base Model provides the decline of IE from early 2000s to later 2000s, then increasing to higher levels in the past decade, after the 2008 financial crisis, while the Alternative Model tends to show continuous growth throughout time, like Matos & Veiga (2019)

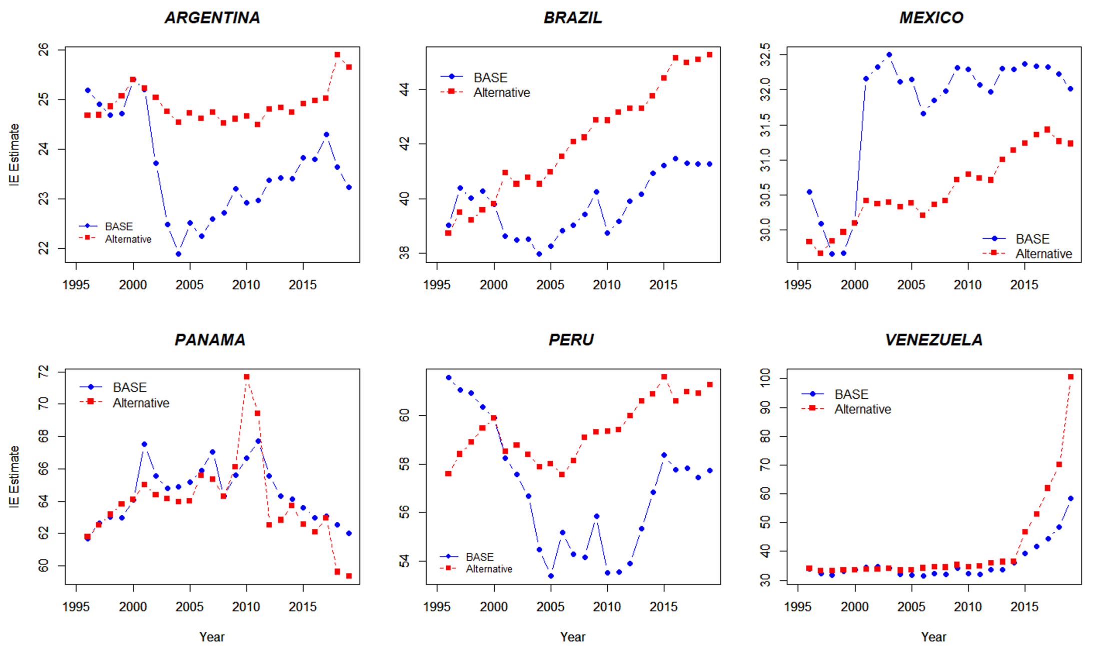


Figure 13 – Alternative Model compared to Base Model

Figure 14 shows the results of the Alternative Model to the different regions in LAC. Here the IE estimate upward trend is easier to spot. With this model we tend to see an upward trend over 20 years in several countries, differently from the Base Model, which we found more consistent with results obtained by other researchers.

While the Base Model is sensitive to GINI and Size of Government (as per Table 5), we see that Taxation and Credit are key to the Alternative Model. Credit is a difficult causal variable to consider in this model, because in theory, the higher the credit the lower IE should be; however, our model showed a positive relation of IE with credit, which can

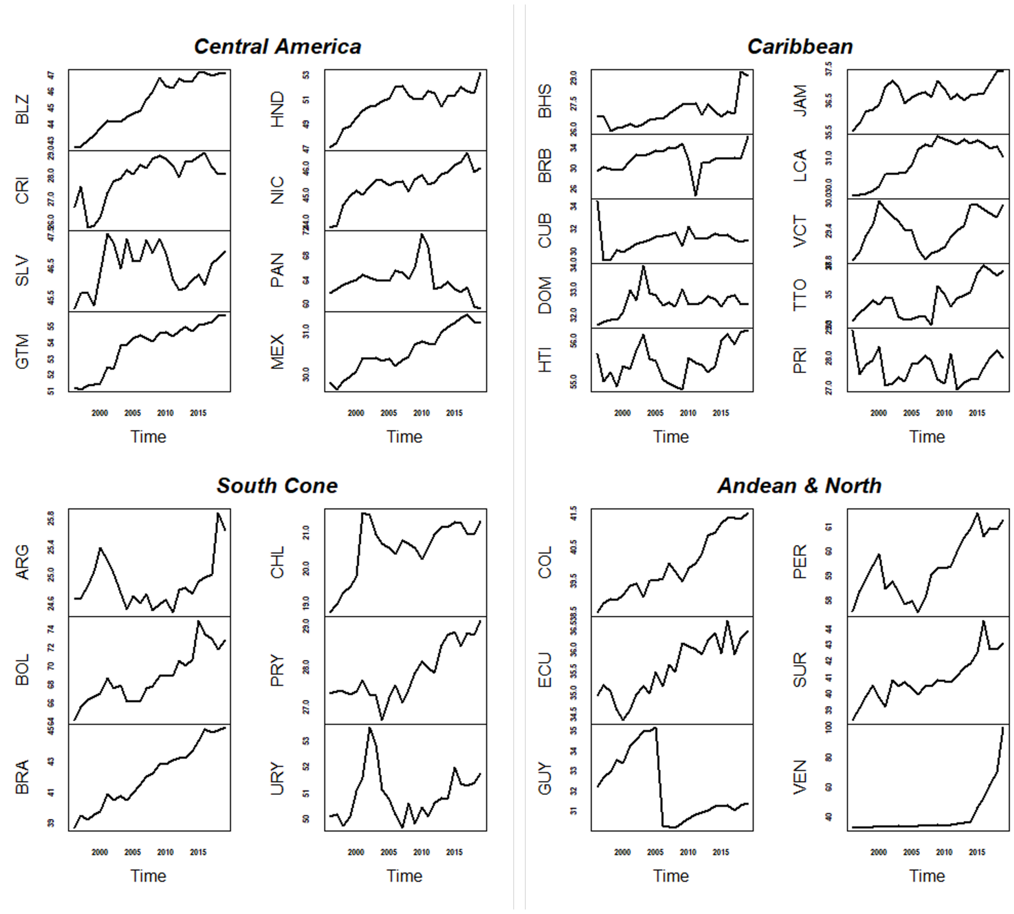


Figure 14 – IE estimates for LAC regions with the Alternative Model

eventually make sense in certain countries in LAC, where entrepreneurship does not choose between being formal or informal.

# C APPENDIX III - DISCUSSION ON GINI VS INFORMAL ECONOMY

Much is needed in terms of further research on analyzing the relationship between inequality and informality. Our Base Model results in a positive impact of inequality on IE. Figure 15 shows IE vs GINI results for 30 LAC countries (6 on the Andean region, 10 in Caribbean, 8 in Central America and 6 in the South Cone), using the average values of IE (estimates resulting from our Base Model) and GINI. We excluded some smaller countries in the Caribbean and other regions from our original sample of 41 countries, to improve the presentation and readability on the graphs (same format was used throughout the article, for both GINI and TFP graphs). The line is a regression between IE fitted values and GINI, for the sample of countries.

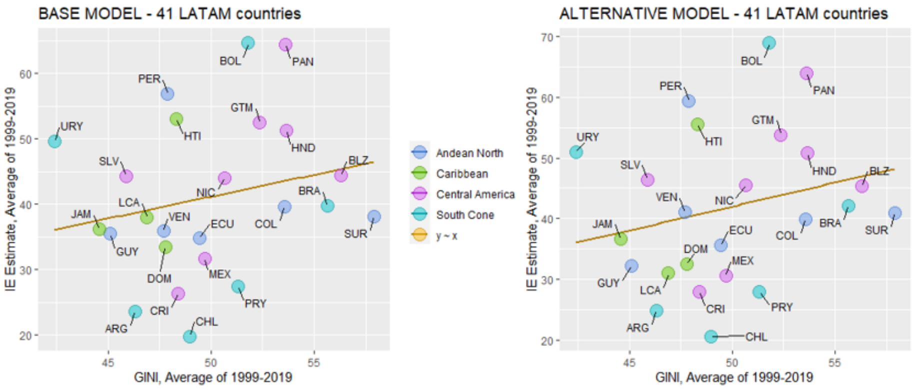


Figure 15 – IE vs GINI results for LAC countries, using our Base Model with data from 41 LAC countries

We exercised other Samples in addition to Appendix II, with the intention to better understand how GINI would behave if the data sets were more focused on developing nations, or with data sets where countries had historically higher levels of IE.

Figure 16 shows the comparison between different Samples.

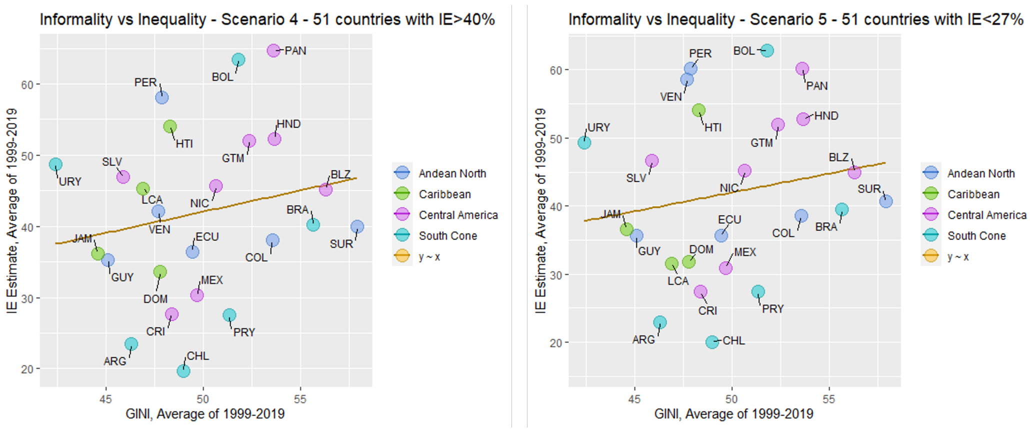


Figure 16 – IE vs GINI results for Sample 4 with IE>40% and Sample 5 with IE<27%

Sample in the left is the Base Model built with data from 51 countries with IE>40%, and Sample in the right with data from 51 countries with IE<27%. Then we run IE estimates for the same 30 countries in LAC, and we plot the results in the same way we did in Figure 15.

IE estimates vary a little, with exception to Venezuela. The estimated value with the model using data from countries with IE<27% is in the range of 60%, like Peru, while the model with IE>40% provides estimated IE in the range of 40%, more in line with our Base Model run with data from 41 LAC countries.

The regression behavior is slightly different too. We see higher inclination on IE>40%, which can indicate that a change in GINI can lead to higher changes in IE. It may indicate that regions like LAC, where IE is high, can benefit from improvements in inequality resulting in greater improvements on the formalization of the economy. On countries with lower IE, the same improvements in inequality result in more limited improvements on the formalization.

Figure 17 shows a comparison between data from developed countries and emerging markets and developing economies. Again, when we plot the estimated IE values for 30 LAC countries, results are similar. Venezuela does not show a huge disparity such as obtained in Figure 16.

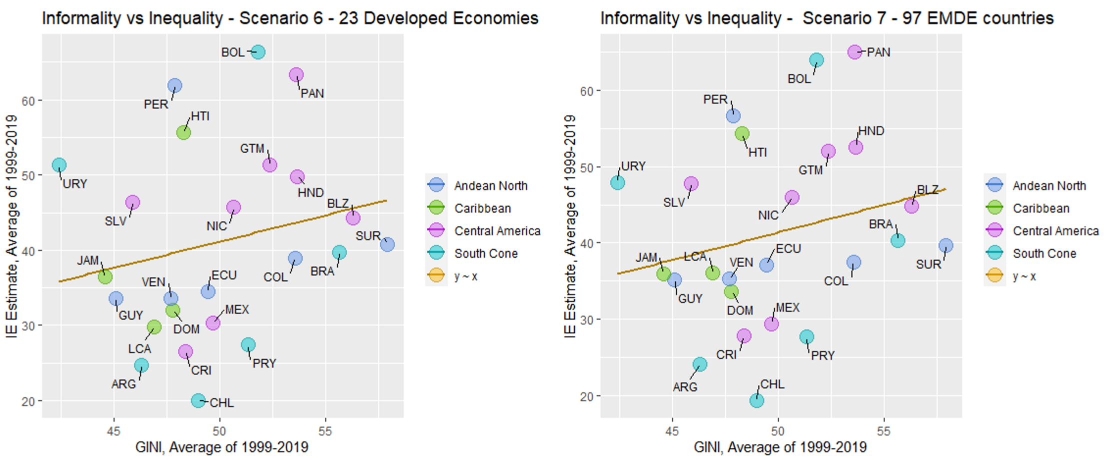


Figure 17 – IE vs GINI results for Sample with Developed countries and EMDE countries

All in all, improvements on inequality reduce the shadow economy.