Inequality, persistence of the informal economy, and club convergence

CGR Working Paper 103

Roxana Gutiérrez-Romero

Abstract

Several new dualistic models have re-examined the causes of the informal economy and have made testable predictions about the long-lasting role of inequality. We test these predictions using historical indicators of inequality, dating back to the 1700s, and data on the informal economy across 138 countries over the 1991–2015 period. We find that past levels of inequality are the most salient factors explaining the size of the informal economy, while improving credit access, reducing tax burden and business costs play a minor role. These results are robust to using alternative inequality measures from various years during the 1700–1992 period, using instrumental variables, and four alternative measurements of the informal economy. Moreover, there is no evidence that the informal economy is converging to the same steady state. Instead, there is convincing evidence of club convergence. Countries with the highest levels of initial inequality are diverging from those that started with lower levels of inequality and those who have made significant redistribution. Results suggest the importance of early conditions in determining the persistence of the dual economy with important policy implications.

Version – April 2020

CGR WORKING PAPER SERIES

Inequality, persistence of the informal economy, and club convergence

Roxana Gutiérrez-Romero[♦]

This version: April 2020

Abstract

Several new dualistic models have re-examined the causes of the informal economy and have made testable predictions about the long-lasting role of inequality. We test these predictions using historical indicators of inequality, dating back to the 1700s, and data on the informal economy across 138 countries over the 1991–2015 period. We find that past levels of inequality are the most salient factors explaining the size of the informal economy, while improving credit access, reducing tax burden and business costs play a minor role. These results are robust to using alternative inequality measures from various years during the 1700–1992 period, using instrumental variables, and four alternative measurements of the informal economy. Moreover, there is no evidence that the informal economy is converging to the same steady state. Instead, there is convincing evidence of club convergence. Countries with the highest levels of initial inequality are diverging from those that started with lower levels of inequality and those who have made significant redistribution. Results suggest the importance of early conditions in determining the persistence of the dual economy with important policy implications.

Keywords Informal economy, Inequality, Club convergence, Institutions

JEL Classification Codes E2, D63, O47, O5

---

1 Introduction

Neoclassic dual economy models, such as the Lewis model, once predicted that the traditional informal sector would be absorbed by the jobs created by the modern economy in the long-run (Lewis, 1954). Yet, the modern economy, intensive in capital, has been unable to absorb the informal labour force. For instance, the informal economy still employs 50-90% of the non-agricultural labour force in the developing world and up to a third in more industrialised economies.[1] To explain the persistence of the informal economy, several new dualistic models have been proposed.[2] These new theoretical models predict that initial levels of inequality determine the size of the informal economy over time. Further, a dangerous permanent dualism is more likely to occur in economies with high initial inequality and inefficient credit markets, contributing to economic divergence across regions and with important consequences for development. Stubbornly high levels of informality imply that as economies modernise, some depending on their initial circumstances, will continue having a significant fraction of businesses evading taxes. Lower public finances, in turn, will limit the ability of the state to cater for disadvantaged groups, most of them employed in the informal sector under precarious conditions (Singer, 1970; Tokman, 2007). Despite the relevance of these new dualistic model predictions, empirically, little is known about the long-term effect of inequality on informality which is critical to understand whether the size of the informal

---

[1] The informal economy accounts for 19% of the Gross Domestic Product (GDP) in OECD countries and 30% in transition countries (Schneider & Enste, 2013; Vanek, Chen, Carré, Heintz, & Hussmanns, 2014).

[2] See for instance: Araujo and Rodrigues (2016), Besley, Burchardi and Ghatak (2012), Chong and Gradstein (2007), Dell'Anno (2018), Docquier, Müller and Naval (2014), Gutiérrez-Romero (2007), Mishra and Ray (2010) and Rosser, Rosser and Ahmed (2003).

economy will eventually decline and converge across countries and the required policies needed to accelerate the formalisation of the economy.

This paper makes three important contributions to the empirical literature on the informal economy. First, the paper is the first one to examine whether countries' inequality in the distant past affects the size of the informal economy in the long run. The goal is not to test whether recent inequality fuels the size of the informal economy, or whether informality affects inequality, both critical issues that have been explored extensively (e.g. Chong & Gradstein, 2007; Dell'Anno, 2016a; Rosser, Rosser, & Ahmed, 2000). Instead, our goal is to examine whether inequality in the distant past, going as far back as the 1700s, 1800s, 1900s, and early 1990s has a long-run effect on the size of the informal economy. Second, the paper contributes to the vivid debate on the most effective policies required to accelerate the formalisation of the economy, is it business regulation or redistributive policies? Third, this paper is also the first one to empirically examine whether the informal economy is converging across countries or whether there are signs of a dangerous permanent dualism as new dualistic models predict.

To analyse our research questions, we combine a comprehensive panel containing the size of the informal economy across 138 countries over the 1991–2015 period with historical indicators of inequality dating back to the year 1700 up to the 1990s. The primary data sources used for the income distribution at the country level are of Bourguignon and Morrisson (2002) and Morrisson and Murtin (2011). In terms of the informal economy, our primary data are the estimates of Hassan and Schneider (2016), given the robustness of their method, geographical and long-period coverage.[3]

---

[3] These authors estimate the informal economy using a type of structural equation model, called the Multiple Indicators, Multiple Causes (MIMIC) method.

The paper offers four key findings. First, our results strongly support the theoretical prediction that initial inequality has long-lasting effects on the size of the informal economy. For instance, a 1% increase in the Gini index for the year 1700 leads to a 3% increase in the size of the informal economy over time. We also show that our results remain consistent in sign, magnitude and statistically significant if we use instead Gini indices for several alternative years during the 1700–1992 period. Our results also remain consistent if we use instead the Gini indices for the 1980s taken from one the World Income Inequality Database gathered by UNU-WIDER, one of the most extensive datasets on inequality. Moreover, our results prove to be consistent to using various alternative estimates of the informal economy, such as those estimated by Medina and Schneider's (2018), Hassan and Schneider (2016), as well as the largest panel on the informal economy gathered by the World Bank Enterprise Survey (WBES) and the International Labour Organisation. These two surveys estimate the percentage of firms that do not pay taxes and the percentage of people employed in the informal economy.

Our second key finding reveals that improvements to business environment (such as improving credit access and cutting tax burden and businesses costs) have had a small-to-negligible impact on the size of the informal economy. Since policymakers wishing to affect the size of the informal economy could also affect business regulation, we demonstrate that our findings remain robust to using instrumental variables (IV). As external instruments, we consider whether the country was colonised and the origin of their legal tradition since these strongly influence the quality of institutions over time (La Porta, Lopez-De-Silanes, Shleifer, & Vishny, 1997). The IV estimates again show a positive relationship between inequality and the size of the informal economy.

We also examine whether inequality might affect the informal economy indirectly by shaping inefficient business regulation, a possibility discussed in related literature (Dabla-

Norris, Kochhar, Ricka, Suphaphiphat, & Tsounta, 2015; Savoia, Easaw, & McKay, 2010). Our third key finding reveals that societies that were highly unequal in the year 1700 have worse access to credit, higher business costs, and tax rates today. However, using structural equation modelling, we demonstrate that inequality primarily influences the informal economy directly as the business environment has little to no impact on informality.

We test for convergence using the method developed by Phillips and Sul (2007), which unlike other similar methods, the researcher does not specify based on some prior knowledge or assumptions, which countries are likely to form each club nor how many clubs there might be in the sample. Our fourth key finding shows that the size of the informal economy is not converging to the same steady-state over time. Instead, there is evidence of club convergence. For instance, although in most highly industrialised countries the informal economy is converging to the same low levels, there are other industrialised countries such as Italy, Spain and Thailand that remain with high levels of informal economy. These findings support theoretical predictions that informality will not necessarily decline as economies modernise. Also, supporting the predictions of the new dualistic models, our findings reveal that countries that are converging to higher levels of informality had high levels of inequality in the distant past (e.g. 1700s and 1800s) and continue to have high levels of inequality as in much of Africa and Latin America. Overall, our results inform the debate about the relative importance of redistributive policy and the business environment needed to formalise the economy.

2 Inequality and informality

To explain the persistence of the informal economy, several new dualistic models have been proposed. These models assume that the formal sector has modern technology, is highly productive and requires the payment of taxes and fixed costs for registering with the

authorities. [4] Thus, only those with sufficient wealth can join this sector by either fully paying the relevant costs or borrowing in the formal credit market. Due to credit market imperfections, formal lenders require collateral. The credit-constrained population can, nonetheless, join the informal economy, a low-productive sector that avoids taxes, government regulations and offers lower net benefits than the formal sector. [5]

This new class of dualistic models concludes that initial inequality, the wealth gap between those who can and cannot join the formal sector, is what determines the size of the informal economy in the short and long run (e.g. Besley, Burchardi, & Ghatak, 2012; Chong & Gradstein, 2007; Dell'Anno, 2018). [6] The wealthiest people will have incentives to remain in the formal sector if the costs involved, such as taxes and interest rates in the formal credit market, are not too high. The poorest segment of the population also will have incentives to migrate to the formal sector to increase their rents, but they will only do so if they can afford

---

[4] This new class of dualistic models include Besley, Burchardi and Ghatak (2012), Chong and Gradstein (2007), Dell'Anno (2018), Dessy and Pallage (2003), Loayza (1999), Massenot and Straub (2016), Rauch (1991), Sarte (2000) and Straub (2005).

[5] In these new models the formal sector charges taxes to cover public goods, pensions or health insurance (Dessy & Pallage, 2003). People employed in the formal sector are also typically assumed to earn higher net returns (salaries plus protection benefits) than those in the informal sector.

[6] This new generation of theoretical studies that model explicitly the informal economy are similar to Banerjee and Newman (1993) and Galor and Zeira (1993) where credit market imperfections restrict investments (into entrepreneurship or human capital) of a sector of the population. Both these papers display multiple equilibria over time depending on initial conditions.

the interest rates and collateral required by formal lenders. Some of these new dualistic models show that the informal economy could gradually decline (Ihrig & Moe, 2004). However, most models also show that under certain conditions, both informality and inequality can remain sizeable over time, where a large share of the population will remain employed in the informal sector under precarious conditions (e.g. earning low salaries and with no social security such as a pension). This prolonged dualism will effectively trap a substantial part of the population in informal employment where informal workers will be unable to accumulate enough capital (or human capital) to formalise their activities, there will not be enough formal jobs around (Dell'Anno, 2016b). Permanent dualism is more likely to occur in economies with high initial inequality and inefficient credit markets (e.g. Araujo & Rodrigues, 2016; Besley et al., 2012; Chong & Gradstein, 2007; Dell'Anno, 2018; Gutiérrez-Romero, 2007; Mishra & Ray, 2010; Rosser, Rosser, & Ahmed, 2003).[7] This permanent dualism will result in overall low output and growth and will contribute to the divergence across economies with different initial conditions (Docquier, Müller, & Naval, 2014).

Despite the importance of these long-run theoretical predictions, the empirical literature has not empirically tested whether inequality in the long-run affects the size of the informal economy nor whether there is evidence of a permanent dualism or the informal economy is declining and converging across economies. Nonetheless, a few empirical studies have explored whether there is any long-term relationship between the level of development and informality. Perhaps the most comprehensive analysis is the work of Wu and Schneider

---

[7] These findings are in line with a broader class of growth theories that show economies similar in structural characteristics (such as institutions and technology) converge to different steady state equilibria if their initial conditions differ (Azariadis & Drazen, 1990; Banerjee & Newman, 1998, 1998; Galor, 1996).

(2019) who find a U-shaped relationship between the size of the informal economy and GDP per capita. Their findings suggest that the informal economy does not necessarily decline as economies industrialise. These findings are in line with those of Baklouti and Boujelbene (2020) who using structural equation modelling, find that informality does not necessarily decline with increases in GDP per capita in countries with poor institutional quality.

Also, a few studies in the empirical literature have explored the short-term and likely endogenous association between inequality and the size of the informal economy.[8] In this respect, some studies have found a positive association for developing countries and a null or negative relationship for economies in transition (Dell'Anno, 2016a; Yap, Sarmidi, Shaari, & Said, 2018).[9] These mixed findings suggest that perhaps there might be a non-linear relationship between inequality and informality, at least in the short-run. Dell'Anno (2018) explains that countries with extremely low or high levels of inequality both can have high levels of informality. For instance, more egalitarian countries might have high levels of informality (in the form of evasion of taxes) if there is excessive redistribution, which makes

---

[8] For instance see Araujo and Rodrigues (2016), Chong and Gradstein (2007), Dell'Anno (2018), Elveren and Özgür (2016), Mishra and Ray (2010), Rosser et al. (2000), and Valentini (2009).

[9] These short-term associations can have several explanations. For instance, redistributive policies might enable some informal entrepreneurs to formalise their activities over time. At the same time, intense competition between formal and informal businesses might affect the gap in returns between these sectors, thereby affecting inequality (Eilat & Zinnes, 2002).

less attractive working in the formal economy.[10] Highly unequal countries are also likely to end up with high levels of informality as inequality increases the financial costs associated with formal business activity. Although this rationale is plausible, it seems that again it is initial conditions, particularly inequality the one affecting the size of the informal economy (directly or indirectly).[11]

Since taxation and business regulation are likely to impact investments, a large body of the empirical literature has instead focused on testing whether efficient government regulation could reduce the size of the informal sector, as proposed by Hernando de Soto. He suggests that informality could be reduced if governments to stop charging high fees and taxes to formalise businesses and poor people have better access to credit markets (De Soto, 1989, 2000).[12] However, despite the progress made over the last two decades in microfinance

---

[10] Several studies before have also suggested that inequality can lead societies to vote for inefficient institutions, including excessively high business regulation and inefficient tax rates (e.g. Alesina & Rodrik, 1994; Benabou, 1996; Savoia et al., 2010).

[11] Dell'Anno (2016b) makes similar rational to explain why highly developed countries might also have high levels of informality. He argues that if these countries impose excessive taxation and business regulation, informality, in the form of tax evasion will raise. Developing countries, can also have high levels of informality (but in the form of informal production) driven by lower level of capital (and human capital) endowment as well as inefficient regulation. Given that the nature of informality activity might be quite varied it is therefore important to explore the relationship between inequality and informality using different data sources as well as to explore the role of business regulation and taxation.

[12] This literature has analysed whether the informal economy reacts to changes in tax rates (Loayza, 1999), property rights (Besley et al., 2012; Johnson, Kaufmann, & Zoido-Lobaton,

and the reduction of bureaucratic red tape, the informal economy has remained stubbornly large (World Bank, 2019).

3 Data

3.1 Measuring the informal economy

The informal economy includes businesses and employment relationships that offer unlicensed services without paying taxes and social benefits (Chen, 2012). Over a dozen methods have been proposed to measure the extent of such informal activities (e.g. Breusch, 2016; Kirchgaessner, 2016; Schneider, 2017; Tanzi, 1999). All these proposed methods, without exceptions, have advantages and disadvantages. Thus, the most recommended approach is to rely on different estimations to validate the size and trends of the informal economy (Schneider & Enste, 2013). In this vein, we use four alternative estimates of the informal economy.

Our primary source is Medina and Schneider's (2018) estimates of the informal economy for 158 countries over the 1991–2015 period. These authors estimate the informal economy using the Multiple Indicators, Multiple Causes (MIMIC) method. The MIMIC method assumes that the informal economy is a latent variable, an unobserved phenomenon, due to known and observable causes of illicit behaviour which can indirectly be observed via some macro-indicators. This method estimates the informal economy using a structural equation model which simultaneously analyses the exogenous factors known to cause the

---

1998), microcredit (Bauchet, Marshall, Starita, Thomas, & Yalouris, 2011; Besley et al., 2012), labour regulation (Maloney, 2004), regulation of business entry (Bruhn, 2008; Djankov, La Porta, Lopez-de-Silanes, & Shleifer, 2002) and the quality of government regulation (Chen, 2007; Perry et al., 2007).

informal economy and the effects the informal economy has on other macro-indicators such as the demand for currency and the employment rate.

We use Medina and Schneider's (2018) estimates as they offer several advantages. Their estimates have by far the largest geographical reach and periods covered. These estimates have also been widely used and cited in the empirical and theoretical literature of informality (e.g. Baklouti & Boujelbene, 2020; Colombo, Menna, & Tirelli, 2019; Huynh & Nguyen, 2019). But, more importantly, Medina and Schneider (2018) offer methodologically improved estimates upon previous MIMIC specifications by measuring the overall economic activity based on satellite data on night lights instead of GDP. In this way, they addressed the main concerns with earlier MIMIC estimates which used GDP both as an exogenous causal factor of the informal economy and as one of the indicators being affected by the informal economy.

In Section 5, we re-run all our econometric specifications using three alternative measures of the informal economy. The first alternative measure is Hassan and Schneider's (2016) estimates of the informal economy across 151 countries over the 1999–2013 period. These authors also employed the MIMIC method but measuring economic activity using the GDP, which might bias the specifications.[13] Since these and similar previous MIMIC

---

[13] The two MIMIC sources used of the informal economy have key differences. For instance, as external factors, Hassan and Schneider (2016) use the GDP growth rate, the labour force participation rate and the currency circulating in the economy, M1/M2. Medina and Schneider (2018) use these external factors as well, but they capture the overall economic activity based on satellite data on night lights instead of GDP. In terms of exogenous factors, Hassan and Schneider (2016) use the total tax revenues as percentage of GDP; government spending; unemployment rate, self-employment rate; indices of economic and business

estimates have also been widely used in the literature (Ball, Furceri, Leigh, & Loungani, 2019; Kirchgaessner, 2006), it is essential and relevant to assess whether our results would differ if using earlier MIMIC estimates. As shown in Figure 1, Medina and Schneider's (2018) estimates turn slightly smaller with a more marked downward trajectory than those of Hassan and Schneider (2016). Despite these differences, as shown later on, our conclusions remain robust regardless of which MIMIC estimates are used.

As a second alternative measure, we use the percentage of the labour force employed in the informal sector across 39 countries over the 2004–2016 period, as estimated by the International Labour Organisation (ILO) using labour surveys. Lastly, we use the percentage of businesses that do not report sales for tax purposes across 73 countries over the 2002–2009 period, as estimated by the World Bank Enterprise Survey (WBES). These two alternative specifications have the main advantage of using direct measures, surveys, to estimate the size of the informal economy. Their disadvantage of these estimates is that the sample is much reduced, but as shown later, these alternative estimates show that inequality in the distant past is positively associated to the informal economy regardless of which estimates of the informal economy is used and period analysed.

---

freedom. In contrast, Medina and Schneider (2018) use as exogenous factors the trade openness; unemployment rate; size of government; rule of law, control of corruption; government stability, and an index of fiscal freedom that measures direct and indirect taxation at all levels of government.
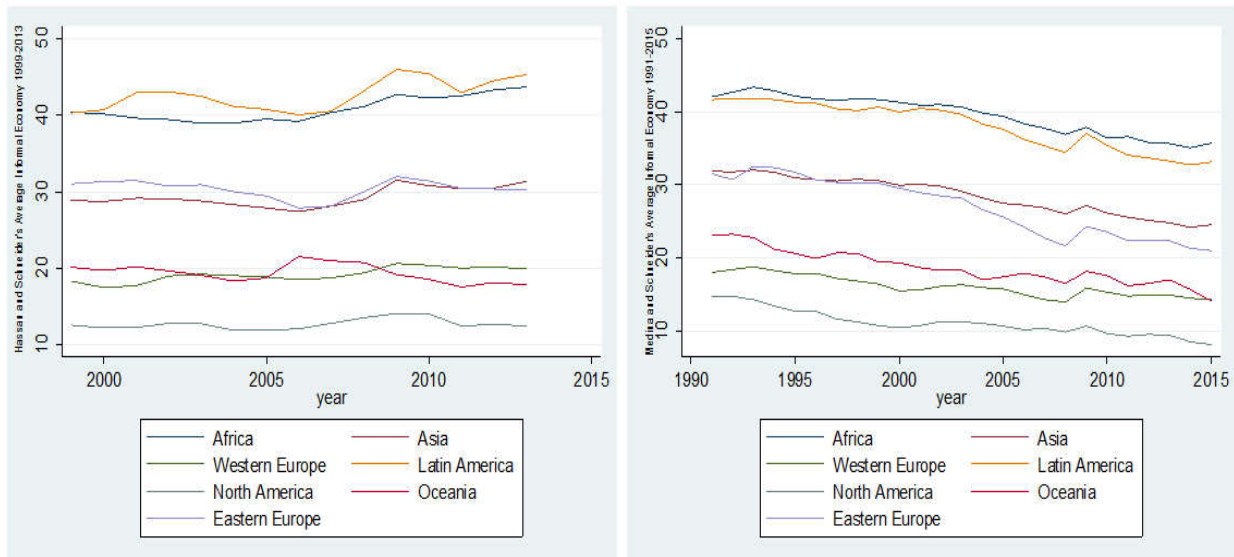
Panel A   Panel B

Fig. 1 Average informal economy by regions.

Panel A uses the informal economy estimates by Hassan and Schneider (2016). Panel B uses

the informal economy estimates by Medina and Schneider (2018).

3.2 Historical indicators of inequality

Since there are scant historical indicators of wealth, we use income inequality as a proxy. We

analyse the effects of the Gini index, a widely used measure of inequality, for 138 countries

estimated at various years during the 1700–1992 period. We get these Gini indices from the

income distributions that Bourguignon and Morrisson (2002) estimated at the country level at

various intervals over the 1820–1992 period.[14] These authors estimated the income

distributions at the country-level for countries with significant large populations. For smaller

countries, the income distribution was estimated in small subgroups of similar economic

---

[14] Table A.1 shows the formula used to interpolate the Gini indices using the income shares

by deciles at country-level.

evolution and history. The full list of countries and sub-group of countries used by Bourguignon and Morrisson (2002) can be found in Table A.1.[15] We complement this dataset with Morrisson and Murtin's (2011) country-level estimates of the income distribution for the year 1700.[16]

We acknowledge that using historical indicators of inequality going as far back as the 1700s and 1800s can be subject to a large margin of error, thus are only indicative for the potential association between inequality and the informal sector. To test the robustness of this association, we also include more recent indicators of inequality up to the year 1990 also estimated by Bourguignon & Morrisson (2002). In addition, we test whether our results remain consistent if we use instead the Gini index taken from the World Income Inequality Database (WIID), gathered by UNU-WIDER. WIDD is one of the most comprehensive sources of both historical and current statistics of income inequality, nonetheless, it has the important caveat that it includes only a small sample of countries for historical inequality indicators. For instance, it has less than 50 countries for the early 1980s and an even much-

---

[15] It is worth noting that our results remain robust if we use exclusively the countries for which there are individual estimates of inequality or whether we remove some or all of the group of countries that their inequality levels were estimated by sub-groups, such as those in Africa, Asia, Latin America, see results in section 5.6.

[16] Bourguignon and Morrisson (2002) estimated the world income distribution mostly based on Maddison's historical record of countries' real GDP and population size. The world income distribution was then later estimated for year 1700 by Morrisson and Murtin (2011). These authors based their estimation on the internal income distributions that Bourguignon and Morrisson (2002) had earlier estimated for the year 1820, also using the same list of countries and subgroups.

reduced sample for the early 1960s. Despite the differences in samples, our results remain consistent.

3.3 Business environment

We also consider business environment indicators mentioned in the literature as important drivers of the informal economy (Antunes & Cavalcanti, 2007; Massenot & Straub, 2015; Srinivas, 2016).[17] We include countries' total tax rate payable by businesses as a share of commercial profits (available during 2006–2013), the cost of opening a business (2004–2013), annual lending interest rates (1995–2015), and an index measuring the ease of getting credit (2005–2013). This index is based on the strength of legal rights and the depth of credit information, thus serving as an alternative proxy for efficiency in the credit market. Table A.1 provides further details about the data, including time period and sources used.

3.4 Summary statistics

Table 1 summarises the average size of the informal economy for the countries for which we have historical indicators of inequality. This table also shows the average Gini index for several years during the 1700–1992 period and our main country-level controls. According to Hassan and Schneider's estimates, the average size of the informal economy over 1999–2013 was 33% of the GDP, a similar figure if using Medina and Schneider's estimates (31%).

Table A.2 provides regional averages that reveal important differences. Latin America and Africa have the largest informal economies, with this sector contributing over 40% of their GDP. Next in the ranking are Asia and Eastern Europe, with their informal economies accounting for 30% of their GDP. Western Europe, Oceania, and North America have the lowest share of the informal economy, contributing less than 20% of their GDP.

---

[17] For a comprehensive review of this literature see Chen (2012), Schneider and Enste (2013) and La Porta and Shleifer (2014).

For an initial graphic examination, Figure 2 shows there is a general positive association between the Gini index for the year 1700 and the average size of the informal economy at the country level. This positive pattern is observed regardless of the MIMIC estimates used, and it becomes even stronger if using more recent indicators of inequality instead, such as for the year 1992. In the next section, we assess whether this association is robust to adding other controls and using random effects model and instrumental variable specifications.

Table 1 Summary statistics

		All sample			
	Number of countries	Mean	SD	Min	Max
Gini index for year 1700 (Morrisson and Murtin, 2011)	138	0.441	0.077	0.3356	0.6204
Gini index for year 1870 (Bourguignon and Morrisson, 2002)	138	0.442	0.077	0.3356	0.6204
Gini index for year 1929 (Bourguignon and Morrisson, 2002)	138	0.448	0.067	0.3544	0.6204
Gini index for year 1960 (Bourguignon and Morrisson, 2002)	138	0.432	0.078	0.2766	0.6204
Gini index for year 1992 (Bourguignon and Morrisson, 2002)	138	0.415	0.093	0.2324	0.58
Gini index for year circa 1980 (UNU-WIDER)	44	0.362	0.111	0.205	0.632
% Change in Gini indices for years 1700 and 1992	138	-0.05	0.178	-0.476577	0.412803
Informal economy 1999-2013 (Hassan and Schneider, 2016)	136	33.21	14.32	9.09	72.30
Informal economy 1991-2015 (Medina and Schneider, 2018)	138	31.38	12.69	7.24	62.28
Employed in Informal Economy 2004-2016	39	54.83	20.71	12.4	84.14999
Firms do not pay sales tax 2002-2009	73	51.14	19.56	12.95	97.32
Lending interest rate 1999-2013	106	16.77	15.69	3.274107	118.2169
Getting credit score 2005-2013	137	49.6	22.83	6.25	100
Cost of starting a business (% of income per capita) 2004-2013	137	63.68	107.3	0.049774	786.5212
Total tax rate (% of profit) 2006-2013	137	49.26	37.45	11.27877	259.6036
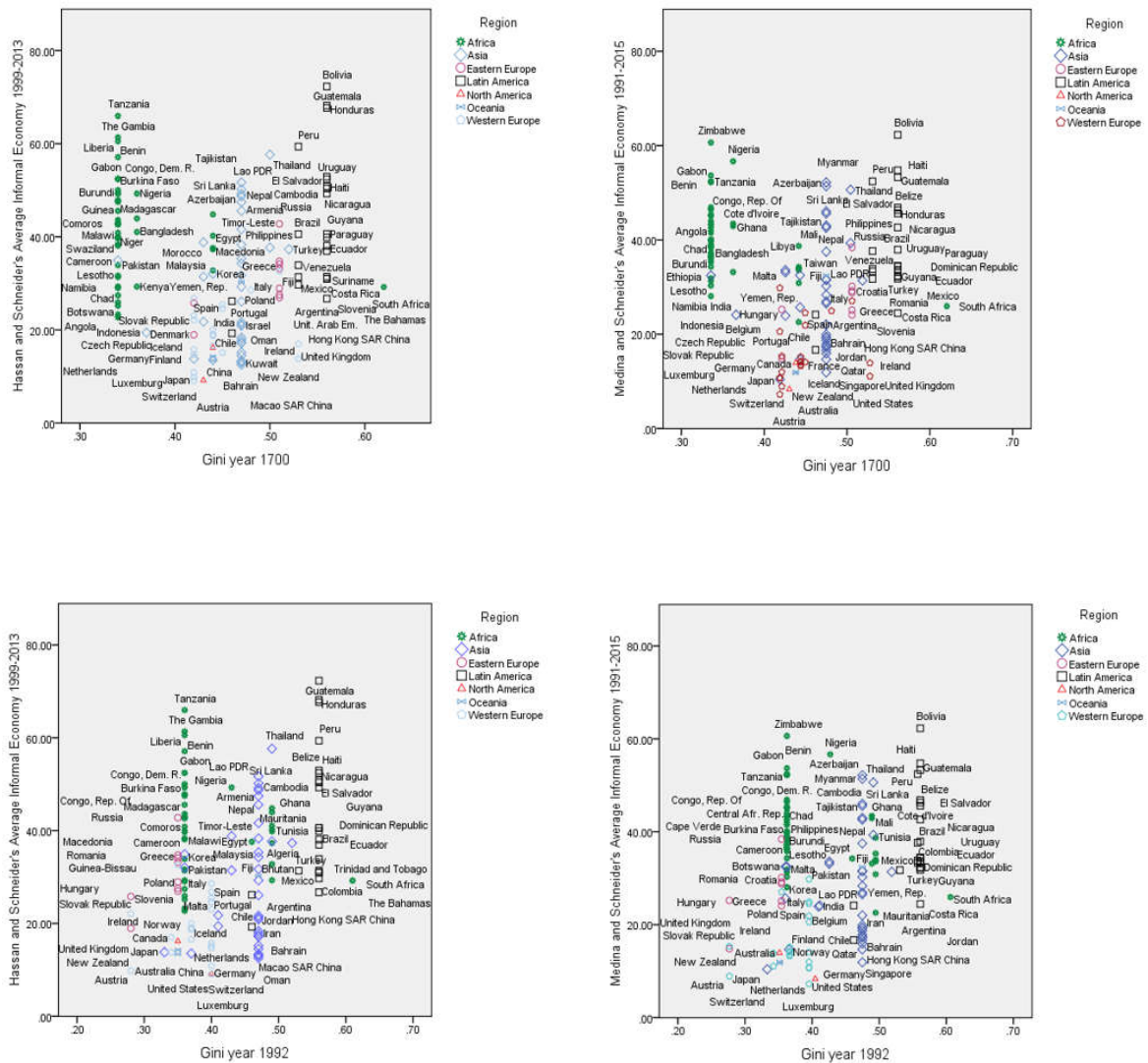
Fig. 2 Informal economy and inequality

4 Does inequality affect the informal economy in the long run?

To examine the association between inequality and the size of the informal economy, we use a random effects (RE) model. The RE model is also known as the multilevel, hierarchical or mixed model since it deals with hierarchical data (in our case having repeated observations nested within countries, the higher-level fixed units) and has the flexibility of simultaneously modelling both time-variant and time-invariant effects (Bell & Jones, 2015). That is, RE model can estimate both time-variant characteristics and different country-level fixed

effects.[18] As shown in Equation (1), with this RE model we estimate the effect of inequality in the distant past, and more recent changes in business regulation on the size of the informal economy.

$$Log(y_{it}) = \beta_0 + \beta_1 LogGini_i + \beta_2 \Delta Gini_i + \beta_3 \mu_i + \beta_4 z_t + \beta_5 X_{it} + \beta_6 \Delta Inst_{it} + (\eta_i + \varepsilon_{it}) \qquad (1)$$

where the dependent variable is $Log(y_{it})$, the natural logarithm of the size of the informal economy (as a percentage of GDP) for country i and year t. In the time-invariant part of the model, $LogGini$ denotes the country's inequality for the year 1700 (or alternative year) in logarithm form. $\Delta Gini$ stands for the percentage change in inequality that the country had between the years 1700 and 1992. μ_i and z_t denote in which country, continent and year the observation corresponds to. Some specifications also include country-level controls that are time-variant X_{it} which include countries' lending interest rate, the score of getting credit, the minimum cost of opening a business, and the total tax rate as a percentage of business profits.

---

[18] Time-invariant characteristics help us to quantify to what extent differences in the informal economy are product of differences in country-specific features, such as socio-economic structures, culture or crucially past levels of inequality. These country effects, along other time-variant characteristics, can be estimated in the RE model provided that there are enough degrees of freedom, as it is our case (Bryan & Jenkins, 2016). These RE estimates will be unbiased if there are no strong sources of endogeneity such as omitted variable due to unobserved heterogeneity. Other methods, such as the fixed effects (FE) model avoids this bias by taking a first difference of the data which removes any time invariant components of the model. Provided that the unobserved heterogeneity is constant over time, the FE model estimates are unbiased, but at the expense of being unable to estimate the effect of time-invariant characteristics. To address any potential bias concern with our RE estimates section 5.2 we use instead instrumental variables.

As time-variant, we also include in some specifications *ΔInst* which measures the change in a specific institutional variable between periods *t* and *t-1*. The so-called 'random' part of the model in parenthesis consists of the residual η_i for country *i*, which allows for differential intercepts for countries (the higher-level entities), and ε_{it} which stands for the time-varying residuals.

4.1 Benchmark results

Table 2 shows our main results. The dependent variable is Medina and Schneider's estimates of the informal economy. In column (1) we begin by including as main regressors the coefficient of the *LogGini* index for the year 1700, the year and country-fixed effects only. This first specification shows that a 1% increase in the Gini coefficient leads to a statistically significant increase of 5.3% in the size of the informal economy.

In column 2, we introduce regional fixed effects, lending interest rates, the cost of starting a business, and the total tax rate. Since we allow these country-level controls to vary over time, there is a risk of endogeneity, which we address using instrumental variables in Section 5.2. Bearing in mind this potential limitation, the RE effect specification shows the regression coefficient of the Gini index remains statistically significant and positive, albeit of a smaller magnitude (2.36). In column 3, we replace the lending rate for the score of getting credit, while in column 4, we add these two variables simultaneously. These alternative specifications show, once again, that the Gini index for the year 1700 is positive and statistically significant. The magnitude of this coefficient is reduced considerably if adding the lending interest rate since the number of countries with this information is reduced (columns 2, and 4). Nonetheless, the regression coefficients of the lending interest rate, score of getting credit, cost of starting a business, and tax burden all have regression coefficients equal to zero, showing no meaningful association with the size of the informal economy.

In columns 5-6, we add the percentual change in inequality, as measured by the Gini index, that countries experienced between the years 1700 and 1992. The results suggest that a 1% increase in inequality between the years 1700 and 1992 is associated with a further increase in the informal economy by less than 1%. In columns 7-8, we also consider whether any recent institutional improvements affect the size of the informal economy. Thus, we add the change between time t and t-1 experienced by the lending interest rate, the score of getting credit, cost of starting a business, and total tax rate. When adding the change in these institutional variables, the regression coefficient of the Gini index remains positive and statistically significant. However, the results suggest that the institutional improvements considered are not associated with a substantial reduction in the size of the informal economy.

Table 2 Historic inequality and Medina and Schneider's informal economy, random effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Log Informal Economy				
LogGini index for year 1700	5.299***	2.361***	5.987***	2.262***	2.373***	3.990***	2.104***	3.305***
	(0.000)	(0.049)	(0.093)	(0.149)	(0.053)	(0.111)	(0.054)	(0.006)
% Change in Gini indices for years 1700 and 1992					0.082*	0.657***	-0.091**	0.937***
					(0.045)	(0.061)	(0.041)	(0.004)
Lending interest rate		-0.000***			-0.000***	-0.000***		
		(0.000)			(0.000)	(0.000)		
Score to get credit			-0.001	-0.000		-0.001		
			(0.001)	(0.001)		(0.001)		
Cost of starting a business (% of income per capita)		0.000**	0.000*	0.000**	0.000**	0.000*		
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)		
Total tax rate (% of profit)		0.000	0.000	0.000	0.000	0.000		
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)		
ΔLending interest rate $_{t\text{-}t\text{-}1}$							-0.002**	
							(0.001)	
ΔGetting credit score $_{t\text{-}t\text{-}1}$							-0.000	-0.000
							(0.000)	(0.000)
ΔCost of starting a business (% of income per capita) $_{t\text{-}t\text{-}1}$								-0.001*
								(0.000)
ΔTotal tax rate (% of profit) $_{t\text{-}t\text{-}1}$							0.000	0.000
							(0.000)	(0.000)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,450	816	1,117	739	816	1,117	719	980
Number of countries	138	95	137	95	95	137	94	126

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

5 Does inequality affect the informal economy? Robustness tests

While our benchmark results suggest inequality in the distant past has a long-term impact on the size of the informal economy, a number of questions remain. In this section, we investigate to what extent this relationship is robust to using different samples, inequality indices referring to other periods, instrumental variables, alternative estimates of the informal economy, alternative data source of inequality, adding a quadratic term of inequality and controlling for differences in the level of development. We also explore whether inequality affects the informal economy indirectly by shaping the business environment.

5.1 More recent inequality indicators

We re-run all our specifications using the Gini index for years 1870, 1929, 1960, or 1992 (all measured in natural logarithm). This test is essential as more recent Gini indices are likely to have a smaller error of measurement than more distant indicators. As Table 3 shows, the association between these alternative Gini indices and the informal economy remains positive, statistically significant, and of a similar magnitude to the one found earlier (presented in Table 2). The change in the Gini indices between the years 1700 and 1992 remains positive and statistically significant for most specifications.[19] Another notable finding reflected in Table 3 is that, again, the regression coefficients of the business environment indicators are zero of very small.

---

[19] For some of the specifications where the percentual change of inequality has a negative sign, such as in columns 15-20, note that the Gini index of 1992 is included twice. That is, where the Gini index is included directly as a regressor and also when estimating the percentual change in inequality between 1700 and 1992. For these cases, the regression coefficient of the change in inequality between 1700 and 1992 becomes positive if removing from the regression the Gini index for the year 1992, result not shown in Table 3.

5.2 Instrumental variable specifications

The specifications thus far shown might suffer from endogeneity. For instance, countries wishing to reduce the size of their informal economies could reduce taxes or costs to business, yielding our institutional controls as endogenous. Thus, as second robustness check, we use instrumental variables. The potentially endogenous variables considered are the score of getting credit, cost of starting a business in the formal sector, and total tax rate for businesses. As external instruments, we use the origin of countries' legal codes and whether they were former colonies.[20] We use these instruments to represent current tax capacity, and the rule of law largely depends on the quality of institutions implemented by former colonisers (La Porta, Lopez-de-Silanes, & Shleifer, 2008; La Porta et al., 1997; Messaouda & Teheni, 2014).

---

[20] We have five binary variables depicting the colonial origin, whether the country was colonised by Great Britain, France, Spain, Portugal or was never colonised by a western power. We also use binary instruments depicting whether the legal code is from French, socialist or common law origin.

Table 3 Alternative inequality indices and Medina and Schneider's informal economy, random effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
										Log Informal Economy										
LogGini index for year 1700	2.373***	3.990***	3.305***																	
	(0.053)	(0.111)	(0.006)																	
LogGini index for year 1870				2.104***	3.990***	3.305***														
				(0.054)	(0.111)	(0.006)														
LogGini index for year 1929							2.352***	2.086***	3.996***	3.270***										
							(0.053)	(0.054)	(0.111)	(0.006)										
LogGini index for year 1960											1.918***	1.701***	4.059***	2.245***						
											(0.043)	(0.044)	(0.113)	(0.004)						
LogGini index for year 1992															2.273***	2.016***	3.998***	3.330***	1.940***	3.189***
															(0.051)	(0.052)	(0.111)	(0.006)	(0.136)	(0.006)
% Change in Gini indices for years 1700 and 1992	0.082*	0.657***	0.937***	-0.091**	0.657***	0.937***	-0.101**	-0.254***	0.272***	0.635***	1.589***	1.245***	-3.386***	-0.938***	-3.279***	-3.072***	-3.832***	-2.808***	-1.817***	-3.563***
	(0.045)	(0.061)	(0.004)	(0.041)	(0.061)	(0.004)	(0.043)	(0.037)	(0.071)	(0.004)	(0.072)	(0.075)	(0.171)	(0.001)	(0.057)	(0.037)	(0.183)	(0.002)	(0.200)	(0.011)
Lending interest rate	-0.000***						-0.000***				-0.000***				-0.000***					
	(0.000)						(0.000)				(0.000)				(0.000)					
Getting credit score		-0.001			-0.001				-0.001				-0.001				-0.001		-0.000	
		(0.001)			(0.001)				(0.001)				(0.001)				(0.001)		(0.001)	
Cost of starting a business (% of income per capita)	0.000**	0.000*			0.000*		0.000**		0.000*		0.000**		0.000*		0.000**		0.000*		0.000	
	(0.000)	(0.000)			(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	
Total tax rate (% of profit)	0.000	0.000			0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.001	
	(0.000)	(0.000)			(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.002)	
ΔLending interest rate $_{t\text{-}t\text{-}1}$				-0.002**				-0.002**				-0.002**				-0.002**				
				(0.001)				(0.001)				(0.001)				(0.001)				
ΔGetting credit score $_{t\text{-}t\text{-}1}$			-0.001*			-0.001*				-0.001*				-0.001*				-0.001*		-0.000
			(0.000)			(0.000)				(0.000)				(0.000)				(0.000)		(0.001)
ΔCost of starting a business (% of income per capita) $_{t\text{-}t\text{-}1}$			-0.000	-0.000		-0.000		-0.000		-0.000		-0.000		-0.000		-0.000		-0.000		0.000
			(0.000)	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)
ΔTotal tax rate (% of profit) $_{t\text{-}t\text{-}1}$			0.000	0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.000		-0.000
			(0.000)	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	816	1,117	980	719	1,117	980	816	719	1,117	980	816	719	1,117	980	816	719	1,117	980	726	633
Number of countries	95	137	126	94	137	126	95	94	137	126	95	94	137	126	95	94	137	126	93	82

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4 IV Estimates of inequality indices and Medina and Schneider's informal economy

	(1) Model 1	(2) Model 2	(3) Model 1	(4) Model 2	(5) Model 1	(6) Model 2	(7) Model 1	(8) Model 2	(9) Model 1	(10) Model 2
					Log Informal Economy					
LogGini index for year 1700	0.965* (0.532)	0.813* (0.469)								
LogGini index for year 1870			1.142* (0.596)	0.881* (0.500)						
LogGini index for year 1929					1.261* (0.665)	0.916* (0.555)				
LogGini index for year 1960							0.900** (0.412)	0.833* (0.482)		
LogGini index for year 1992									1.001*** (0.342)	0.719* (0.436)
% Change in Gini indices for years 1700 and 1992		1.325*** (0.295)		1.313*** (0.293)		1.171*** (0.329)		1.160*** (0.350)		0.489 (0.645)
Getting credit score	-0.004 (0.004)	-0.003 (0.004)	-0.003 (0.005)	-0.002 (0.004)	-0.003 (0.005)	-0.001 (0.004)	-0.002 (0.004)	-0.000 (0.004)	-0.003 (0.005)	-0.002 (0.004)
Cost of starting a business (% of income per capita)	0.006* (0.003)	0.003 (0.003)	0.006* (0.003)	0.003 (0.003)	0.006* (0.003)	0.003 (0.003)	0.005** (0.002)	0.003 (0.003)	0.004* (0.003)	0.003 (0.003)
Total tax rate (% of profit)	-0.002 (0.006)	0.002 (0.004)	-0.002 (0.006)	0.002 (0.004)	-0.002 (0.006)	0.002 (0.004)	0.000 (0.005)	0.004 (0.005)	0.001 (0.006)	0.003 (0.004)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region2 fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,117	1,117	1,117	1,117	1,117	1,117	1,117	1,117	1,117	1,117
Number of countries	137	137	137	137	137	137	137	137	137	137
Hausman test										
Chi2	3.600	6.370	2.540	5.170	2.690	5.160	4.920	4.440	6.930	6.570
Prob>chi2	0.995	0.956	0.999	0.983	0.997	0.984	0.977	0.992	0.906	0.950

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The first-stage regressions, shown in Table A.3, show that the excluded instruments are strongly associated to the endogenous regressors. Table 4 shows the second-stage specifications, which include two model specifications denoted as Model 1 and Model 2. Both specifications use the same external instruments. However, in Model 1, the size of the informal economy is regressed on the assumed exogenous Gini index and the instrumented institutional variables. In Model 2, we add the assumed exogenous percentual change in the Gini index between the years 1700 and 1992. The results reveal that the effect of the Gini index on the informal economy is positive and statistically significant. These IV results are consistent, regardless of using the Gini indices for the year 1700 or the Gini index for alternative years (e.g. 1870, 1929, 1960 and 1992). Similarly, the regression coefficient of the percentual change between the Gini index for years 1700 and 1992 remains positive and statistically significant in all Model 2 specifications, except for column 10 where the Gini index for 1992 is also included.

Table 4, in the bottom row, shows the results of the Hausman test. This test reveals whether the IV specifications should be preferred instead of the earlier presented RE model. The Hausman test suggests that across all columns (1-10), there are no systematic differences between the IV specifications (Table 4) and the RE model (Table 2).

5.3 Alternative MIMIC estimates of the informal economy

As a third robustness check, we use alternative sources to measure the size of the informal economy. Our first alternative proxy is the informal economy estimates obtained by Hassan and Schneider (2016) for 151 countries for the 1999–2013 period. Table A.4 shows that when using this alternative estimate, the Gini index for the year 1700 has a positive and statistically significant effect on the size of the informal economy. The similarity in results is maintained when using Gini indices for various alternative years (1870, 1929, 1960, 1992), as shown in Table A.5. These alternative Gini indices are again positively associated with the informal

economy, even after controlling for the business environment or using instrumental variables. For instance, Table A.6 shows the second-stage IV regression if using the same instruments as the previous regression, while the first-stage are presented in Table A.7. The IV results show that the Gini index is positively associated with the size of the informal economy, regardless of the Gini index used (for years 1700, 1870, 1929, 1960, or 1992). Additionally, the percentual change for the Gini index between the years 1700 and 1992 is positive and statistically significant.

5.4 Measuring informality using labour surveys

The consistency in results, thus far, might be related given that we have used similar MIMIC estimates of the informal economy. As an additional robustness check, we use a more direct estimate of the informal economy—the percentage of the labour force employed in the informal sector, as estimated by the ILO based on labour surveys. As shown in Table 5, the effect of the Gini index on the labour force employed in the informal economy is positive and statistically significant. As also shown in Table 5, these results are consistent regardless of using the Gini indices for the year 1700 or the Gini index for alternative years (e.g. 1870, 1929, 1960 and 1992). Similarly, the regression coefficient of the percentual change for the Gini index for between 1700 and 1992 stays positive and statistically significant in most specifications. In this analysis, we also consider the score of getting credit, cost of starting a business, and tax rate; none of these indicators are statistically significant.

Since these results might suffer from endogeneity, we re-run our results using instrumental variables. We instrument these potentially endogenous variables using the same instruments as before—whether the country has been a colony and the origin of its legal code. Table A.9 reports the corresponding first-stage regression which shows that the excluded instruments explain the instrumented variables reasonably well. Table A.8 shows the second-stage IV results. The impact of inequality on the share of the labour force employed in the

informal economy remains positive and statistically significant. This result holds true if using the Gini index for year 1700, if using the Gini index for more recent years (e.g. 1870, 1929, 1960, 1992), and if adding the percentual change in inequality that countries experienced between 1700 and 1992. The IV specifications also show that score of getting credit and lower taxes both tend to reduce the share of the population employed in the informal economy, albeit by small margin (-0.06% and -0.18% respectively as shown in column 7).

In Table A.8, we also include alternative specifications. As potentially endogenous variables, we use the change between periods t and t-1 in the lending interest rate, the cost of starting a business in the formal sector, and the total tax rate. For these variables only, we use the following instrumental variables: the lending interest rate in year 1995, the cost of starting a business, and the tax rate, all of which are lagged for one period. These alternative instrumental specifications suggest that inequality has a long-run effect on the informal economy, while improvements in the business environment have no statistically significant impact (columns 3, 6, 9, 12, and 15).

5.5 Measuring informality using business surveys

To conclude with alternative estimates of the informal economy, we use the percentage of businesses that do not report sales for tax purposes. Table 6 presents the RE model specifications using as main controls the Gini index and the change in inequality between 1700 and 1992. The results suggest that the percentual change in the Gini index between 1700 and 1992 is positive and statistically significant. Similar to our earlier results, an increase of 1% in the Gini index leads to approximately a 1% increase in the businesses that do not pay taxes. In this analysis, we also consider the score of getting credit, cost of starting a business, and tax rate; none of these indicators are statistically significant.

Table 5 Inequality and labour force employed in the informal economy as estimated by the International Labour Organisation (ILO) using labour surveys, random effects model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
				Log Employed in Informal Economy						
LogGini index for year 1700	4.710***	4.710***								
	(0.783)	(0.783)								
LogGini index for year 1870			4.710***	4.710***						
			(0.783)	(0.783)						
LogGini index for year 1929					4.210***	4.710***				
					(0.358)	(0.783)				
LogGini index for year 1960							4.210***	4.710***		
							(0.358)	(0.783)		
LogGini index for year 1992									6.868***	4.710***
									(0.585)	(0.783)
% Change in Gini indices for years 1700 and 1992		8.867***		8.867***		-1.409		-1.409		3.724***
		(0.740)		(0.740)		(1.641)		(1.641)		(0.944)
Getting credit score	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Cost of starting a business (% of income per capita)	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Total tax rate (% of profit)	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	90	90	90	90	90	90	90	90	90	90
Number of countries	30	30	30	30	30	30	30	30	30	30

Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Since these results might suffer from endogeneity, we re-run our results using instrumental variables, using same instruments as before. Table A.11 reports the corresponding first-stage regression and Table A.10 shows the second-stage IV results. The impact of inequality on the percentage of businesses that do not report sales for tax purposes remains positive and statistically significant. This result holds if using the Gini index for the year 1700, if using the Gini index for more recent years (e.g. 1870, 1929, 1960, 1992), and if adding the percentual change in inequality that countries experienced between 1700 and 1992.

5.6 Excluding countries for which income distribution was estimated in large groups

Thus far, we have tested the robustness of our results using Gini indices for several years, all taken from the same sources Bourguignon and Morrisson (2002) and Morrisson and Murtin (2011). These indices have two shortcomings. First, using historical indicators undoubtedly will introduce a measurment error. Second, the country-level income inequality was estimated independently for large countries but using sub-group of countries for smaller countries with similar socio-economic characteristics, which might contribute to a larger estimation error for these sub-groups. To address for any potential concern with estimating income distribution in large groups, in Table 7, re-run our specification using the Medina and Schneider's estimates of the informal economy excluding all the countries for which income inequality indicators were estimated in large sub-groups.[21] Our sample is reduced significantly, leaving only 63 countries. But we still find a positive and statistically significant association between the informal economy and past inequality (e.g. the Gini index for 1700 and other more recent indicators such as the Gini index for the year 1992).

---

[21] That is, we remove 46 African countries for which their income-distributions were estimated as a block, as well as other 45 Asian countries and 37 Latin American countries.

Table 6 Firms that do not pay sales taxes as estimated by the World Bank based on business surveys, random effects model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
					Log Firms do not Pay Sales Tax					
LogGini index for year 1700	0.886***	1.410***								
	(0.222)	(0.259)								
LogGini index for year 1870			0.886***	1.410***						
			(0.222)	(0.259)						
LogGini index for year 1929					0.888***	1.222***				
					(0.222)	(0.224)				
LogGini index for year 1960							0.656***	1.042***		
							(0.164)	(0.191)		
LogGini index for year 1992									1.068***	1.367***
									(0.267)	(0.251)
% Change in Gini indices for years 1992 and 1700		3.619***		3.619***		2.295***		3.599***		1.712***
		(0.376)		(0.376)		(0.255)		(0.373)		(0.263)
Getting credit score	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Cost of starting a business (% of income per capita)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Total tax rate (% of profit)	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	299	299	299	299	299	299	299	299	299	299
Number of countries	75	75	75	75	75	75	75	75	75	75

Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

In Table 7, we also test whether the effect of inequality might be quadratic and depend on the level of development, as some empirical studies have suggested (Dell'Anno, 2016b). As shown in columns 4 and 9, we add the square term of the log of Gini as well as the log of GDP for the year 1870.[22] Both these columns show that there is no evidence that inequality has a quadratic effect on the size of the informal economy. In columns 3 and 8 we introduce an interaction between the Gini coefficient and region, revealing that there are no statistically significant differences by region.

5.7 UNU-WIDER Gini indices

Here we re-run our results but using much more recent income inequality indicators, drawn from an alternative data source. We use instead the Gini index taken from the World Income Inequality Database (WIID), gathered by UNU-WIDER. This database offers one of the most comprehensive historical and current statistics of income inequality. Although this database provides an excellent alternative source, there are no more than 50 countries with indicators of income inequality for the year circa 1980 and a much-reduced sample for earlier than 1960. Nonetheless, using this alternative inequality index, we still find the same strong association between inequality and the size of the informal economy if using Medina and Schneider's estimates (see Table 8, columns 1-6).

---

[22] We include such lagged value of GDP to avoid introducing endogeneity and lead to spurios regression by having current GPD in the estimates of the informal economy and as a control as well.

Table 7 Historic inequality and Medina and Schneider's informal economy excluding countries for which Bourguignon and Morrisson estimated income distribution in large groups of countries, random effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Log Average Informal Economy										
LogGini index for year 1700	2.792***	3.284***		2.427	2.584***	3.206***					
	(0.000)	(0.203)		(3.792)	(0.088)	(0.013)					
Squared log Gini index year 1700				0.967							
				(2.265)							
LogGini index for year 1992							3.311***		2.736	2.475***	3.230***
							(0.205)		(1.962)	(0.085)	(0.013)
Squared log Gini index year 1992									0.863		
									(0.891)		
% Change in Gini indices for years 1992 and 1700	0.708***	0.600***			0.050	0.223***	-3.114***			-3.611***	-3.409***
	(0.000)	(0.167)			(0.081)	(0.014)	(0.104)			(0.143)	(0.005)
Getting credit score		0.000	0.000	-0.000			0.000	0.000	0.000		
		(0.001)	(0.001)	(0.001)			(0.001)	(0.001)	(0.001)		
Cost of starting a business (% of income per capita)		0.000	0.000	-0.000	-0.000		0.000	0.000	-0.000	-0.000	
		(0.001)	(0.001)	(0.001)	(0.000)		(0.001)	(0.001)	(0.000)	(0.000)	
Total tax rate (% of profit)		0.001	0.001	0.001	0.003***		0.001	0.001	0.001	0.003***	
		(0.001)	(0.001)	(0.002)	(0.001)		(0.001)	(0.001)	(0.002)	(0.001)	
Log GDP per capita for year 1870				-0.696***					-0.603***		
				(0.161)					(0.158)		
Lending interest rate					0.004					0.004	
					(0.002)					(0.002)	
ΔGetting credit score $_{t-t-1}$						-0.000					-0.000
						(0.001)					(0.001)
ΔCost of starting a business (% of income per capita) $_{t-t-1}$						0.001**					0.001**
						(0.000)					(0.000)
ΔTotal tax rate (% of profit) $_{t-t-1}$						0.001					0.001
						(0.001)					(0.001)
Africa x Log Gini index 1700			0.000								
			(0.000)								
Asiax x Log Gini index 1700			0.000								
			(0.000)								
Western Europe x Log Gini index 1700			0.000								
			(0.000)								
Latin America x Log Gini index 1700			0.000								
			(0.000)								
North America x Log Gini index 1700			0.000								
			(0.000)								
Oceania x Log Gini index 1700			0.000								
			(0.000)								
Eastern Europe x Log Gini index 1700			0.000								
			(0.000)								
Africa x Log Gini index 1992									0.000		
									(0.000)		
Asiax x Log Gini index 1992									0.000		
									(0.000)		
Western Europe x Log Gini index 1992									0.000		
									(0.000)		
Latin America x Log Gini index 1992									0.000		
									(0.000)		
North America x Log Gini index 1992									0.000		
									(0.000)		
Oceania x Log Gini index 1992									0.000		
									(0.000)		
Eastern Europe x Log Gini index 1992									0.000		
									(0.000)		
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes
Observations	1,575	465	465	340	260	402	465	465	340	260	402
Number of countries	63	63	63	44	31	52	63	63	44	31	52

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

One could argue that we find the same positive association between inequality and the informal economy by pure chance, just by focusing on a smaller group of countries. But, as shown in Table 8, columns 7-10, we show that we find the same positive association between inequality and the informal economy if we use instead the Gini index the year 1700 focusing only on the 44 countries for which there are Gini indices gathered by UNU-WIDER.

We also find a positive association between the Gini index for year 1980 gathered by UNU-WIDER and the informal economy estimates obtained by Hassan and Schneider's (Table A.12), the percentage of the labour force employed in the informal economy (Table A.13, column 1), and the percentage of firms that do not pay sales taxes (Table A.13, column 4).

5.8 Testing for a quadratic effect of inequality

As mentioned earlier, some empirical studies have suggested that recent levels of inequality might have a quadratic effect on the size of the informal economy. We find no statistically significant evidence to suggest that inequality has a quadratic effect on the size of the informal economy regardless of whether using the Gini index of UNU-WIDER for the year 1980 or the Gini index for the year 1700 or 1992. The same is the case for Medina and Schneider's estimates (Tables 8, columns 4-5, 9-10, 14-15), Hassan and Schneider's estimates (Tables A.12, columns 6-7, 11-12, 16-17), for labour force employed in the informal economy (Table 9, columns 2-3) and for firms evading sales taxes (Table 9 columns 5-6).

Table 8 Inequality and Medina and Schneider's informal economy only for countries with available WIID Gini index, random effects model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
							Log Informal Economy									
Log WIID Gini index circa year 1980	1.248***	2.469***	2.546***	1.560	5.684*	1.193***										
	(0.165)	(0.000)	(0.439)	(1.455)	(2.908)	(0.104)										
Squared log WIID Gini index circa year 1980				0.180	1.965											
				(0.632)	(1.229)											
% Change in WIID Gini and indice for years 1980 and historical Gini 1700		-6.679***	-6.798***													
		(0.000)	(1.156)													
LogGini index for year 1700							1.086***	1.166***	-0.636	1.822	7.692***					
							(0.000)	(0.209)	(4.931)	(15.391)	(0.672)					
Squared log Gini index year 1700									-0.744	0.666						
									(3.205)	(9.995)						
LogGini index for year 1992												1.087***	1.167***	0.927	2.696	1.437***
												(0.000)	(0.209)	(2.452)	(3.068)	(0.125)
Squared log Gini index year 1992														-0.039	0.838	
														(1.120)	(1.328)	
% Change in Gini indices for years 1992 and 1700							-1.833***	-1.822***				-3.750***	-3.879***			
							(0.000)	(0.302)				(0.000)	(0.670)			
Getting credit score			-0.002***	-0.002***	-0.000	0.001		-0.002***	-0.002***	0.000	0.001		-0.002***	-0.002***	0.000	0.001
			(0.001)	(0.001)	(0.002)	(0.002)		(0.001)	(0.001)	(0.002)	(0.002)		(0.001)	(0.001)	(0.002)	(0.002)
Cost of starting a business (% of income per capita)			0.000*	0.000*	0.000	0.000		0.000*	0.000*	0.001	0.000		0.000*	0.000*	0.001	0.000
			(0.000)	(0.000)	(0.003)	(0.003)		(0.000)	(0.000)	(0.003)	(0.003)		(0.000)	(0.000)	(0.003)	(0.003)
Total tax rate (% of profit)			0.001***	0.001***	0.001	0.000		0.001***	0.001***	0.001	0.000		0.001***	0.001***	0.001	0.000
			(0.000)	(0.000)	(0.002)	(0.002)		(0.000)	(0.000)	(0.002)	(0.002)		(0.000)	(0.000)	(0.002)	(0.002)
Log GDP per capita for year 1870					-0.541***	-1.199***				-0.641***	-3.024***				-0.587***	-1.103***
					(0.194)	(0.160)				(0.199)	(0.292)				(0.187)	(0.155)
Country fixed effects	No	Yes	Yes	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No	Yes
Year fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,100	1,100	355	355	255	255	1,100	355	355	255	255	1,100	355	355	255	255
Number of countrycode	44	44	44	44	31	31	44	44	44	31	31	44	44	44	31	31

Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 9 Alternative WIID Gini index, labour force employed in the informal economy and firms that do not pay sales taxes, random effects

	(1)	(2)	(3)	(4)	(5)	(6)
	Log Employed in Informal Economy			Log Firms do not Pay Sales Tax		
Log WIID Gini index circa year 1980	3.774***	11.286	2.329***	1.023***	0.790	0.617
	(0.020)	(34.133)	(0.000)	(0.193)	(1.690)	(1.681)
Squared log WIID Gini index circa year 1980		7.887	0.000		-0.126	-0.224
		(19.595)	(0.000)		(0.963)	(0.959)
Getting credit score	-0.041***		-0.012	0.000***		0.000***
	(0.000)		(0.000)	(0.000)		(0.000)
Cost of starting a business (% of income per capita)	-0.002		0.017***	0.000		0.000
	(0.000)		(0.000)	(0.000)		(0.000)
Total tax rate (% of profit)	0.028		-	-	-	-
	(0.000)					
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12	23	12	89	144	89
Number of countrycode	4	6	4	18	18	18

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

5.9 Is inequality affecting the informal economy indirectly?

An extensive body of literature shows that wealth concentrated in the hands of a few can lead to creating inefficient institutions, including inefficient taxes and credit markets, as well as high entry costs to protect the business elite (see Alesina & Rodrik, 1994; Benabou, 1996; Dabla-Norris et al., 2015; Savoia et al., 2010). Therefore, inequality might affect the informal economy via both direct and indirect channels. As the dualistic literature suggests, inequality can affect the size of the informal economy directly. For instance, an initial inequitable distribution might allow a small group of the population and their offspring to engage in the formal economy, which will allow them earning higher rents than those with lower levels of wealth who will be trapped in the informal economy. Furthermore, initial inequality might also affect the business environment indirectly by affecting tax rates and business regulation.

To assess the potential indirect, direct and total effects of inequality on the informal economy, we use structural equation (SE) modelling. SE models have a wide range of applications (e.g. confirmatory analysis, path analysis, latent analysis) which among others allow to explicitly model complex relationships among various variables that might be directly observed and those that might be indirectly measured (latent variables). Our goal is not to estimate again the size of the informal economy or inequality. Instead, we take all the variables (inequality, size of the informal sector, and the three institutional factors) as given, as already measured, and we explicitly model their relationship in the system of four-linked equations shown in Equations (2)-(5).

$$Log(y_{it}) = \gamma_0 + \gamma_1\ LogGini_i + \gamma_2 \Delta Gini_i + \gamma_3 \mu_i + \gamma_4\ Credit\ score_{it} + \gamma_5\ Cost\ business_{it} + \gamma_6\ Tax_{it} + \varepsilon_{1it} \quad (2)$$

$$Credit\ score_{it} = \delta_0 + \delta_1\ LogGini_i + \delta_2 \Delta Gini_i + \delta_3 \rho_1 + \varepsilon_{2it} \quad (3)$$

$$Cost\ business_{it} = \delta_4 \gamma_0 + \delta_5 LogGini_i + \delta_6 \Delta Gini_i + \delta_7 \rho_1 + \varepsilon_{3it} \quad (4)$$

$$Tax_{it} = \delta_8 + \delta_9 LogGini_i + \delta_{10} \Delta Gini_i + \delta_{11} \rho_1 + \varepsilon_{4it} \quad (5)$$

Equation (2) regresses the informal economy on the score of getting credit, the cost of starting a business, the total tax rate, the Gini index for the year 1700, *LogGini*, the change in inequality between 1700 and 1992, *ΔGini*, and country fixed effects, μ_i. Equations (3), (4) and (5) regress the score of getting credit, the cost of starting a business, and the total tax rate on the Gini index for the year 1700, the change in inequality between 1700 and 1992, and regional fixed effects, ρ_i. ε_{it} denote the white-robust errors, clustered at country level.

We estimate the four-equation SE model using maximum likelihood (ML) using Medina and Schneider (2018) estimates of the informal economy as it offers the largest sample size for all the business indicators analysed. We run three specifications: using the Gini index for the year 1700, another one using the Gini index for the year 1992, and another one using the Gini index for the year 1980, taken from UNU-WIDER.[23] In Figure 3 we show the path analysis used as well as the corresponding indirect, and total effects of inequality on the size of the informal economy. We depict all main variables in rectangles since we are taking their values as already measured (not as latent variables). Also, as standard in SEM analysis, the measurement error of the observed variables are depicted as small ovals. Table 10 shows the net effect of each of the four-equations, as well as the measures of fit for each of these equations.

---

[23] Whenever using the WIID Gini index we omit adding the percentage change in inequality between year 1700 and 1980 so we omit using earlier estimates, but our results do not change much if adding this term.
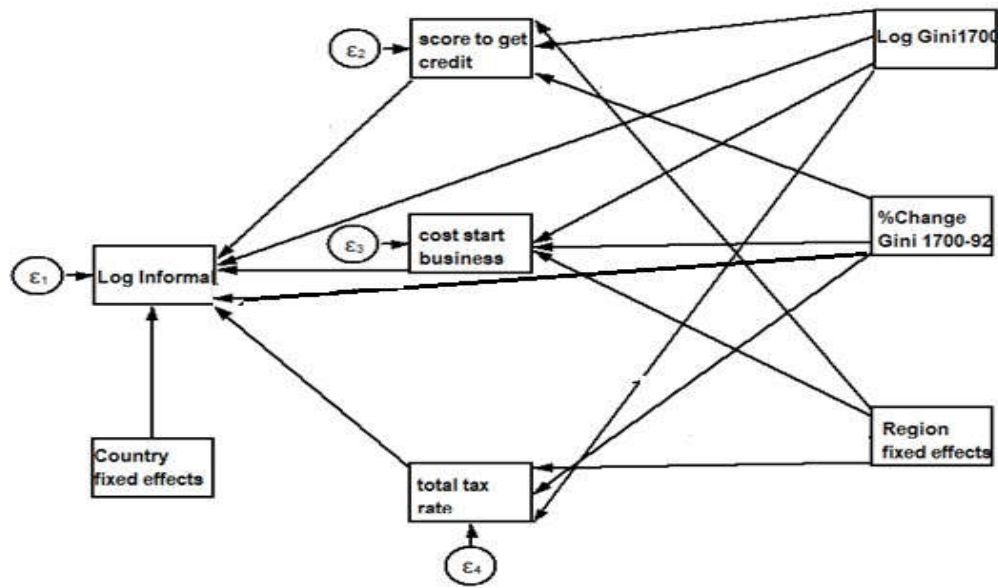
Fig. 3 Structural equation model of inequality, informality and the business environment

As Table 10 shows, inequality in the past, the year 1700, is associated with a higher tax rate and higher costs of starting a business over time but not the ease of getting credit. That is the case if using the Gini coefficient for the year 1700 and 1992 but not the one of WIID, potentially due to differences in samples. Nonetheless, these cotemporaneous business environment indicators have in turn a small-to-negligible impact on the informal economy, as our earlier findings have suggested. Thus, although inequality in the distant past shapes current business institutions, inequality still has a significant direct effect on the size of the informal economy. For instance, Table 10 also shows regardless of which Gini index we use, we find that a 1% increase in the Gini coefficient is associated to a net increase of the informal economy of slightly above 1% (1.7% if using the Gini for the year 1700, 1.5% if using the Gini instead for the year 1992 and 1.005% if using the Gini for 1980 from UNU-WIDER).

Table 10 Structural equation model total effects

	Total Effects		
Dependent variable: Log Informal Economy (Medina and Schneider)			
LogGini index for year 1700	1.753***		
	(0.093)		
LogGini index for year 1992		1.554***	
		(0.081)	1.005***
Log WIID Gini index circa year 1980			(0.054)
% Change in Gini indices for years 1992 and 1700	3.593***	2.102***	
	(0.055)	(0.082)	
Getting credit score	-0.002***	-0.002***	-0.003***
	(0.000)	(0.000)	(0.001)
Cost of starting a business (% of income per capita)	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)
Total tax rate (% of profit)	0.000	0.000	0.001***
	(0.000)	(0.000)	(0.000)
Country fixed effects	Yes	Yes	Yes
R-squared	0.980	0.981	0.984
mc	0.990	0.990	0.992
mc2	0.980	0.981	0.984
Dependent variable: Getting credit score			
LogGini index for year 1700	21.258		
	(21.293)		
LogGini index for year 1992		11.768	
		(20.941)	
Log WIID Gini index circa year 1980			-3.55
			(14.470)
% Change in Gini indices for years 1992 and 1700	-15.800	-31.952	
	(18.192)	(29.158)	
Region fixed effects	Yes	Yes	Yes
R-squared	0.330	0.326	0.311
mc	0.575	0.571	0.557
mc2	0.330	0.326	0.311
Dependent variable: Cost of starting a business (% of income per capita)			
LogGini index for year 1700	-202.007***		
	(59.077)		
LogGini index for year 1992		-168.376***	
		(51.356)	
Log WIID Gini index circa year 1980			19.69
			(30.029)
% Change in Gini indices for years 1992 and 1700	-5.911	194.159***	
	(50.621)	(70.768)	
Region fixed effects	Yes	Yes	Yes
R-squared	0.208	0.204	0.109
mc	0.456	0.451	0.330
mc2	0.208	0.204	0.109
Dependent variable: Total tax rate (% of profit)			
LogGini index for year 1700	-64.628**		
	(26.642)		
LogGini index for year 1992		-49.126**	
		(22.917)	
Log WIID Gini index circa year 1980			-0.223
			(13.380)
% Change in Gini indices for years 1992 and 1700	-14.26	45.842***	
	(17.459)	(25.926)	
Region fixed effects	Yes	Yes	Yes
R-squared	0.111	0.106	0.067
mc	0.333	0.325	0.259
mc2	0.111	0.106	0.067
Overall R-squared	0.987	0.987	0.989

mc is the correlation between the dependent variable and its prediction. mc2 is the Bentler-Raykov squared multiple correlation coefficient. Robust standard errors clustered at country-level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

5.10 Interaction between inequality and business environment

As an additional test, we also analyse whether the effect of inequality is enhanced in countries with worse business environment. To do so, we add interaction terms between the Gini index for various years (1700, 1870, 1929, 1960 and 1992) and the score of getting credit, the cost of starting a business and the total tax rate. In this regression, we use separately the estimates of Hassan and Schneider (2016) and Medina and Schneider (2018). None of these interactions are statistically significant, as shown in Table A.13.

5.11 Does the effect of inequality depend on the level of development?

We also confirm that our results do not depend on countries' level of development. Kirchgaessner (2016) explains that researchers must not include the same controls used in obtaining the MIMIC estimates of the informal economy otherwise, the analysis would yield statistical significant yet trivial results. For this reason, we have not used current countries' GDP as a control in our analysis. Nonetheless, we verified that our results stay robust if adding the GDP per capita prevailing in the 1800s or more recent past. Table A.14, for instance, shows that the positive association between inequality and the informal economy remains robust to adding as control the GDP per capita for the year 1870.[24]

5.12 Running analysis by pooling data in sub-periods

All our earlier RE specifications have allowed us to analyse the long-term association between inequality and the size of the informal economy while controlling country-specific

---

[24] We omit presenting the RE results for the labour force employed in the informal economy and business that do not pay sales tax as their sample is reduced considerably if adding historical measures of GDP, remaining with an overall sample of less than four or nine countries.

changes in the business environment or the informal economy. We also confirm that our conclusions remain consistent if we instead split the sample into 5-10 years sub-periods or collapse the data to the overall averages. For instance, Table 11 shows the pooled OLS results of regressing the average size of the informal economy on the Gini index, the average score of getting credit over 2005–2013, the average cost of starting a business over 2006–2013 and the average total tax rate over 2006–2013. This pooled OLS dataset shows that inequality is positively associated to the average size of the informal economy, regardless of whether using the Gini indices for the year 1700, or the Gini index for alternative years (e.g. 1870, 1929, 1960 and 1992).[25] In these specifications, we use Hassan and Schneider's estimates of the informal economy over 1999–2013 and separately Medina and Schneider's estimates over 1991–2015. In these cross-country specifications, we add the GDP per capita instead for the year 1870 as control. This GDP is distant enough from the MIMIC estimations of the informal economy and a good proxy for the differences in levels of development.

---

[25] The pooled OLS regressions reduce considerably our sample sizes, particularly for the percentage of labour force employed in the informal economy or the business that do not pay sales taxes, reason why we omit presenting these results.

Table 11 Inequality and informal economy, pooled OLS cross-section

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
	Log Average Informal Economy 1999–2013 (Hassan and Schneider)										Log Average Informal Economy 1991-2015 (Medina and Schneider)									
LogGini index for year 1700	1.005*	1.044*									1.079**	1.216***								
	(0.507)	(0.532)									(0.440)	(0.451)								
LogGini index for year 1870			1.357**	1.427**									1.339***	1.471***						
			(0.566)	(0.548)									(0.491)	(0.471)						
LogGini index for year 1929					1.783**	1.825**									1.645***	1.659***				
					(0.731)	(0.739)									(0.589)	(0.583)				
LogGini index for year 1960							1.745***	1.958***									1.354***	1.358***		
							(0.371)	(0.440)									(0.418)	(0.432)		
LogGini index for year 1992									0.347	0.745*									0.623*	0.948**
									(0.377)	(0.413)									(0.342)	(0.362)
% Change in Gini indices for years 1700 and 1992		0.102		0.174		0.106		0.434		-0.908		0.495		0.495		0.228		0.204		-0.694
		(0.531)		(0.531)		(0.566)		(0.338)		(0.637)		(0.475)		(0.471)		(0.474)		(0.358)		(0.500)
Average getting credit score over 2005-2013	0.001	0.001	0.000	0.000	0.001	0.001	0.002	0.002	0.001	0.001	-0.003	-0.003	-0.003	-0.003	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002
	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Average cost of starting a business (% of income per capita) over 2004-2013	0.008**	0.008**	0.008**	0.008**	0.009***	0.009***	0.009***	0.009***	0.008**	0.008**	0.009***	0.008***	0.009***	0.009***	0.009***	0.009***	0.008***	0.008***	0.008***	0.008***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Average total tax rate (% of profit) over 2006-2013	0.007*	0.007*	0.007**	0.007**	0.007**	0.007**	0.008**	0.007**	0.006	0.007*	0.005	0.005*	0.005	0.005*	0.005*	0.005*	0.006*	0.006*	0.004	0.005
	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Log GDP per capita for year 1870	-0.323*	-0.315*	-0.327*	-0.314*	-0.289*	-0.280*	-0.326**	-0.293*	-0.287	-0.331*	-0.407***	-0.386***	-0.414***	-0.393***	-0.381***	-0.370***	-0.421***	-0.411***	-0.369***	-0.394***
	(0.165)	(0.175)	(0.165)	(0.173)	(0.152)	(0.165)	(0.153)	(0.151)	(0.173)	(0.176)	(0.139)	(0.139)	(0.139)	(0.139)	(0.133)	(0.136)	(0.135)	(0.135)	(0.138)	(0.141)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations/number of countries	46	46	46	46	46	46	46	46	46	46	54	54	54	54	54	54	54	54	54	54
R-squared	0.614	0.614	0.635	0.636	0.645	0.646	0.668	0.674	0.580	0.603	0.675	0.684	0.689	0.698	0.684	0.686	0.709	0.711	0.662	0.672

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

6 Testing convergence

In this section, we test the second main prediction of the dual economy models. Is the informal economy converging over time? We probe for convergence using the regression-based test proposed by Phillips and Sul (2007). This regression, known as the log t test, has three main advantages over other traditional convergence tests.[26] First, the log t convergence test analyses whether there is convergence across the whole sample or only among a subset of countries, known as club convergence. Second, the log t test searches for data commonalities using a data-driven algorithm. Therefore, unlike other convergence methods, the researcher does not specify, based on some prior knowledge or assumptions, which countries are likely to form each club nor how many clubs there might be in the sample. Third, the log t test does not suffer from small sample properties, unlike the traditional unit root and cointegration tests which are not suitable for testing convergence in panel data settings like ours.

To analyse the transitional behaviour of the informal economy, the log t regression analyses the evolution of the size of the informal economy. In the appendix, section A.1, we explain all the steps followed by Phillips and Sul (2007) to derive their convergence test. In sum, this method tests the null hypothesis of convergence using a simple regression which includes a ratio of cross-sectional variance H_1/H_t, where H_1 measures the variation at the beginning, of the sample ($t=1$), and H_t represents the variation for every point in time. Taking the log of H_1/H_t, this ratio then measures the distance of the panel from the common limit.

$$\log(H_1/H_t) - 2\log L(t) = \hat{a} + \hat{b} \log t + \widehat{u_t}, \quad \text{for } t=[fT], [fT]+1,....T \text{ with } f>0 \qquad (6)$$

where f denotes the fraction of the initial sample to be removed from the sample before running the regression to alleviate any potential sensitivity of the results to the initial

---

[26] We implement the log t test using Du (2017) logtreg stata package.

conditions. We set f equal to 0.3 as recommended by Phillips and Sul (2007) for samples with reasonable shorter time spells like ours ($T \leq 50$). Phillips and Sul (2007) show the fitted coefficient of the log t regression is $\hat{b} = 2\hat{\alpha}$, where $\hat{\alpha}$ is the estimate of the speed of convergence α. The hypothesis of convergence is rejected if the computed one-sided t-test has a p-value below the stated significance level. For instance, the null hypothesis of convergence is rejected at the 5% significance level if the t-test < -1.65.

Rejecting convergence across the whole panel cannot rule out the existence of convergence across subgroups, or clubs, within the sample. Thus, Phillips and Sul (2007) develop a data-based algorithm that identifies clubs based on the values of the dependent variable, rather than on an imposed restriction. Also, in the appendix, section A.1, we describe in further detail how this algorithm identifies the number of clubs and the respective countries within each club. In sum, this algorithm sorts the series in descending order, then identifies if countries can form a club based on whether their corresponding log t test is larger than -1.65. The log t is repeated for all countries in the sample to check whether they can converge to the same club or a different one. If no clubs are found, then it is concluded that countries are diverging.

6.1 Is the informal economy converging?

We start by testing whether the informal economy is converging across the 138 countries for which we have both the Medina and Schneider's (2018) estimates over 1991–2015 and the historical Gini indices. Separately, as an extra robustness test, we also test for convergence across the 136 countries for which we have information about both the Hassan and Schneider's (2016) estimates of the informal economy over 1999–2013 and the historical

Gini indices.[27]   Table 12 shows that regardless of the informal economy estimates used, we reject the hypothesis of convergence as the t-test < -1.65. That is, there is no evidence the size of the informal economy is converging at the same steady-state equilibrium.[28]

Table 12 Convergence test for the informal economy

	b	SE	t-test
Panel A. Hassan and Schneider (2016)			
log (t)	-1.151	0.005	-224.176
Panel B. Medina and Schneider (2018)			
log (t)	-0.975	0.004	-230.740

6.2 Is there club convergence?

Since in our sample, we included developed and developing countries, it is essential to test whether there is evidence of convergence among certain groups of countries according to their level of similarity in the size of their informal economies. As shown in Table 13, the data-driven algorithm developed by Phillips and Sul (2007) first divides the sample into several clubs, according to their similarity. Table 14 shows that some of these clubs can be

[27] This extra test is relevant since Medina and Schneider's (2018) estimates are more conservative and have a more pronounced decaying trend those of Hassan and Schneider (2016), as shown earlier in Figure 1. There are not enough observations across countries to also test convergence in the percentage of people employed in the informal sector or business avoiding sales taxes, as for most countries we have at most one year of data.

[28] We separately tested for convergence using all countries including those few which we had dropped earlier for lacking historical inequality indicators. We again find no evidence of convergence among all the sample. We omit these non-significant results but are available upon request.

merged. In total, the data can be divided into six distinct club of countries converging, as shown in Table 15.

Table 13 Club convergence for the informal economy

	Club 1	Club 2	Club 3	Club 4	Club 5	Club 6	Club 7	Club 8	Club 9	Club 10	Club 11
Panel A. Hassan and Schneider (2016)											
b	0.092	0.495	0.113	0.427	-0.054	0.28	1.324				
t-test	0.533	3.09	1.198	3.466	-1.154	3.471	7.099				
Panel B. Medina and Schneider (2018)											
b	0.878	0.109	0.971	0.309	0.114	0.482	0.202	1.441	0.23	0.203	0.557
t-test	4.808	2.141	10.628	2.801	1.33	3.898	2.077	-0.747	2.034	3.106	4.111

Table 14 Tests of club merging

	Club 1+2	Club 2+3	Club 3+4	Club 4+5	Club 5+6	Club 6+7	Club 7+8	Club 8+9	Club 9+10	Club 10+11	Club 11+12
Panel A. Hassan and Schneider (2016)											
b	-0.329	0.195	-0.108	-0.39	-0.524	-0.329					
t-test	-12.036	1.988	-1.796	-9.975	-29.081	-20.02					
Panel B. Medina and Schneider (2018)											
b	0.651	0.414	-0.173	-0.208	-0.247	0.091	0.188	0.043	-0.163	-0.078	-1.322
t-test	5.037	11.87	-2.837	-3.184	-3.953	1.025	1.962	0.474	-2.509	-1.532	-88.065

Table 15 Final club selection

	Club 1	Club 2	Club 3	Club 4	Club 5	Club 6
Panel A. Hassan and Schneider (2016)						
b	0.092	0.195	0.427	-0.054	0.280	1.324
t-test	0.533	1.988	3.466	-1.154	3.471	7.099
Informal economy 1999-2013	56.268	41.000	36.053	25.427	17.381	13.343
Gini index for year 1700	0.442	0.440	0.449	0.448	0.439	0.433
Gini index for year 1992	0.454	0.445	0.445	0.390	0.358	0.393
Panel B. Medina and Schneider (2018)						
b	0.316	0.309	0.114	0.013	0.230	-0.078
t-test	3.454	2.801	1.330	0.155	2.034	-1.532
Informal economy 1991-2015	47.055	37.495	31.857	22.393	14.444	11.803
Gini index for year 1700	0.432	0.425	0.442	0.463	0.456	0.446
Gini index for year 1992	0.455	0.434	0.433	0.386	0.395	0.317

In Table 15 we also report the broad characteristics of these clubs, such as their average Gini index[29] and the b coefficient associated with each club. This b coefficient

---

[29] We merely report these average Gini indices to see if there is any correlation between inequality levels and the characteristics of each club. At no point the data-algorithm uses information on inequality to divide the sample into six club of countries.

reveals how fast the members of each club are converging towards the same steady-state. The larger the b coefficient, the faster the convergence. The specific countries contained in each of the clubs are in Table 16 if referring to Medina and Schneider's estimates and in Table A.15 in the appendix if referring to Hassan and Schneider' estimates. Both these tables show that most advanced economies are clustered in the same clubs. For instance, West European countries and the USA are mostly concentrated in either Club 5 or Club 6. The countries that are converging at the fastest rate (according to their b coefficient) have substantial informal economies such as Brazil, Mexico, Pakistan, Venezuela, and several African countries. In general clubs with the larger size of the informal economies had the highest levels of inequality in both 1700 and 1992, supporting new dualistic models' predictions.

Table 16 Club convergence classification by countries using Medina and Schneider's estimates of the informal economy

Club 1 (26 countries)					
Peru	Thailand	Myanmar	The Gambia	Guinea	Cote d'Ivoire
Benin	Liberia	Nigeria	Central Afr. Rep.	Gabon	Congo, Dem. R.
Guinea-Bissau	Tanzania	Burundi	Madagascar	Comoros	Belize
Guatemala	El Salvador	Honduras	Nicaragua	Haiti	Bolivia
Azerbaijan	The Bahamas				
Club 2 (26 countries)					
Egypt	Greece	Italy	Mexico	Brazil	Venezuela
Pakistan	Sri Lanka	Libya	Senegal	Niger	Togo
Ghana	Equatorial Guinea	Congo, Rep. Of.	Angola	Kenya	Zambia
Malawi	Swaziland	Armenia	Paraguay	Cambodia	Yemen
Nepal	Tajikistan				
Club 3 (39 countries)					
Russia	Spain	Romania	Argentina	Colombia	Malasia
Indonesia	Philippines	Turkey	Morocco	Algeria	Tunisia
Mali	Burkina Faso	Sierra Leone	Chad	Cameroon	Cape Verde
Rwanda	Ethiopia	Uganda	Mozambique	Lesotho	Malta
Croatia	Costa Rica	Guyana	Ecuador	Fiji	Lao PDR
Bangladesh	Taiwan	Lebanon	Kuwait	Oman	Unit. Arab. Em.
Bahrain	Dominican Republic	Trinidad and Tobago			
Club 4 (25 countries)					
South Africa	Belgium	Hungary	Denmark	Norway	Poland
Korea	Vietnam	India	Iran	Mauritania	Namibia
Botswana	Portugal	Luxemburg	Iceland	Finland	Bulgaria
Slovenia	Suriname	Uruguay	Jordan	Saudi Arabia	Israel
Bhutan					
Club 5 (8 countries)					
France	Austria	Sweden	Chile	China	Hong Kong SAR China
Qatar	Mongolia				
Club 6 (13 countries)					
United States	Netherlands	Switzerland	United Kingdom	Germany	Australia
New Zealand	Singapore	Japan	Canada	Ireland	Czech Republic
Slovak Republic					
Not convergent Group 7 (1 country)					
Zimbabwe					

7 Conclusion

Neoclassical models predict that as economies industrialise the traditional informal sector would be absorbed by the jobs created by the modern economy as a natural by-product of development (Lewis, 1954). These traditional neoclassical models were written during the 1950s, back when the rapid growth experienced worldwide seem to support the idea that industrialisation would bring a rapid formalisation of the economy, rise in living standards, rule of law and fiscal capacity. Today, it is clear that a significant percentage of the population is still working outside the rule of law, evading taxes, without access to legal or social benefits. To understand this reality, new dualistic models have been proposed (e.g. Besley et al., 2012; Chong & Gradstein, 2007; Dell'Anno, 2018; Rosser et al., 2003). These new models reach two critical predictions: countries with larger initial inequality have over time larger informal economies, and it is not necessarily the case that the informal economy will naturally decline to the same steady state. To the best of our knowledge, this article is the first to empirically test these predictions.

We analysed a comprehensive panel on the informal economy across 138 countries over 1991–2015 and historical inequality indicators dating to 1700s. Our results reveal that inequality is positively and strongly associated with the size of the informal economy in the long run. Our findings are robust to using different inequality indices during the 1800s and 1900s, four alternative measures of the informal economy, business environment, instrumental variables, testing for quadratic effects of inequality, controlling for different levels of development and using structural equation modelling.

Much of the empirical scholarship has argued that inefficient government regulation is the main contributing factor for the still large informal economy. Our study does not dispute that optimal taxation or adequate access to credit is essential for growth. Instead, our results suggest that deregulation or improving credit access have a small impact on the

informal economy. Our empirical findings are in line with the experimental literature that suggests deregulating helps new business to be created in the formal sector but has a small or null impact on existing informal businesses (de Mel, McKenzie, & Woodruff, 2008; La Porta & Shleifer, 2014). Our findings are also consistent with the experimental literature that suggests microcredit has a modest impact on long-term business creation (Banerjee, Karlan, & Zinman, 2015).

We also found that the informal economy is not converging to a common steady-state, but rather to various steady-states across various subgroups of countries. Countries that had high levels of inequality in the distant past are converging to having high contemporaneous levels of informal economy and inequality. These findings have important implications. The informal economy sector has lower productivity, lower output and lower net returns (including social benefits) than the formal economy sector. Thus, countries with a large informal economy will deviate even further from those already industrialised countries.

The overall evidence points to discredit the idea that the informal economy can naturally decline simply by fine-tuning regulation. Instead, as the new dualistic models suggest to see a significant decline in the informal economy, the long-lasting effects of inequality need to be addressed first.

7 References

Alesina, A., & Rodrik, D. (1994). Distributive politics and economic growth. *The Quarterly Journal of Economics*, *109*(2), 465–490.

Alesina, A., & Rodrik, D. (1994). Distributive Politics and Economic Growth. *The Quarterly Journal of Economics*, *109*(2), 465–490. https://doi.org/10.2307/2118470

Antunes, A. R., & Cavalcanti, T. V. de V. (2007). Start up costs, limited enforcement, and the hidden economy. *European Economic Review*, *51*(1), 203–224. https://doi.org/10.1016/J.EUROECOREV.2005.11.008

Araujo, J. P., & Rodrigues, M. (2016). Taxation, credit constraints and the informal economy. *Economia*, *17*(1), 43–55. https://doi.org/10.1016/J.ECON.2016.03.003

Azariadis, C., & Drazen, A. (1990). Threshold externalities in economic eevelopment. *The Quarterly Journal of Economics*, *105*(2), 501–526. https://doi.org/10.2307/2937797

Baklouti, N., & Boujelbene, Y. (2020). A simultaneous equation model of economic growth and shadow economy: Is there a difference between the developed and developing countries? *Economic Change and Restructuring*, *53*(1), 151–170. https://doi.org/10.1007/s10644-018-9235-8

Ball, L., Furceri, D., Leigh, D., & Loungani, P. (2019). Does One Law Fit All? Cross-Country Evidence on Okun's Law. *Open Economies Review*, *30*(5), 841–874. https://doi.org/10.1007/s11079-019-09549-3

Banerjee, A., Karlan, D., & Zinman, J. (2015). Six Randomized Evaluations of Microcredit: Introduction and Further Steps. *American Economic Journal: Applied Economics*, *7*(1), 1–21. https://doi.org/10.1257/app.20140287

Banerjee, A. V., & Newman, A. F. (1993). Occupational Choice and the Process of Development. *Journal of Political Economy*, *101*(2), 274–298. https://doi.org/10.1086/261876

Banerjee, A. V., & Newman, A. F. (1998). Information, the Dual Economy, and Development. *The Review of Economic Studies*, *65*(4), 631–653. https://doi.org/10.1111/1467-937X.00062

Bauchet, J., Marshall, C., Starita, L., Thomas, J., & Yalouris, A. (2011). *Latest findings from randomized evaluations of microfinance* (Access to Finance Forum).

Bell, A., & Jones, K. (2015). Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data. *Political Science Research and Methods*, *3*(1), 133–153. https://doi.org/10.1017/psrm.2014.7

Benabou, R. (1996). Inequality and growth. *NBER Macroeconomics Annual*, *11*, 11–74.

Besley, T. J., Burchardi, K. B., & Ghatak, M. (2012). Incentives and the De Soto Effect. *The Quarterly Journal of Economics*, *127*(1), 237–282. https://doi.org/10.1093/qje/qjr056

Bourguignon, F., & Morrisson, C. (2002). Inequality Among World Citizens: 1820–1992. *American Economic Review*, *92*(4), 727–744. https://doi.org/10.1257/00028280260344443

Breusch, T. (2016). Estimating the Underground Economy using MIMIC models. *Journal of Tax Administration*, *2*(1).

Bruhn, M. (2008). *License to Sell: The Effect of Business Registration Reform on Entrepreneurial Activity in Mexico* (World Bank Policy Research Working Paper Series).

Bryan, M. L., & Jenkins, S. P. (2016). Multilevel Modelling of Country Effects: A Cautionary Tale. *European Sociological Review*, *32*(1), 3–22. https://doi.org/10.1093/esr/jcv059

Chen, M. A. (2007). *Rethinking the Informal Economy: Linkages with the Formal Economy and the Formal Regulatory Environment* (WIEGO Working Papers Series). United Nations, Department of Economics and Social Affairs. Retrieved from

https://ideas.repec.org/p/une/wpaper/46.html

Chen, M. A. (2012). *The informal economy: Definitions, theories and policies* (WIEGO

Working Paper Series).

Chong, A., & Gradstein, M. (2007). Inequality and informality. *Journal of Public Economics*,

91(1–2), 159–179. https://doi.org/10.1016/J.JPUBECO.2006.08.001

Colombo, E., Menna, L., & Tirelli, P. (2019). Informality and the labor market effects of

financial crises. *World Development*, *119*, 1–22.

https://doi.org/10.1016/j.worlddev.2019.03.001

Dabla-Norris, E., Kochhar, K., Ricka, F., Suphaphiphat, N., & Tsounta, E. (2015). *Causes*

and Consequences of Income Inequality : A Global Perspective (IMF Staff Discussion

Note 15/13).

de Mel, S., McKenzie, D., & Woodruff, C. (2008). *Who Are the Microenterprise Owners?*

Evidence from Sri Lanka on Tokman v. De Soto. (World Bank Policy Research Working

Paper 4635).

De Soto, H. (1989). *The other path: The economic answer to terrorism.* New York:

HarperCollins.

De Soto, H. (2000). *The mystery of capital: Why capitalism triumphs in the west and fails*

everywhere else. New York: Basic Books.

Dell'Anno, R. (2016a). Analyzing the Determinants of the Shadow Economy With a

"Separate Approach". An Application of the Relationship Between Inequality and the

Shadow Economy. *World Development*, *84*, 342–356.

https://doi.org/10.1016/J.WORLDDEV.2015.08.026

Dell'Anno, R. (2016b). *Inequality and informality in transition and emerging countries* (IZA

World of Labor).

Dell'Anno, R. (2018). Inequality, informality and credit market imperfections.

Macroeconomic Dynamics, *22*(5), 1184–1206.

https://doi.org/10.1017/S1365100516000663

Dessy, S., & Pallage, S. (2003). Taxes, inequality and the size of the informal sector. *Journal of Development Economics*, *70*(1), 225–233. https://doi.org/10.1016/S0304-3878(02)00086-X

Djankov, S., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2002). The Regulation of Entry. *The Quarterly Journal of Economics*, *117*(1), 1–37. https://doi.org/10.1162/003355302753399436

Docquier, F., Müller, T., & Naval, J. (2014). *Informality and Long-run Growth* (IZA Discussion Paper No. 7883).

Du, K. (2017). Econometric Convergence Test and Club Clustering Using Stata. *The Stata Journal: Promoting Communications on Statistics and Stata*, *17*(4), 882–900. https://doi.org/10.1177/1536867X1801700407

Eilat, Y., & Zinnes, C. (2002). The Shadow Economy in Transition Countries: Friend or Foe? A Policy Perspective. *World Development*, *30*(7), 1233–1254. https://doi.org/10.1016/S0305-750X(02)00036-0

Elveren, A., & Özgür, G. (2016). The effect of informal economy on income inequality: Evidence from Turkey. *Panoeconomicus*, *63*(3), 293–312. https://doi.org/10.2298/PAN1603293E

Galor, O. (1996). Convergence? Inferences from Theoretical Models. *The Economic Journal*, *106*(437), 1056–1069. https://doi.org/10.2307/2235378

Galor, O., & Zeira, J. (1993). Income Distribution and Macroeconomics. *The Review of Economic Studies*, *60*(1), 35. https://doi.org/10.2307/2297811

Gutiérrez-Romero, R. (2007). *The Effects of Inequality on the Dynamics of the Informal Economy* (IZA/WB conference, Bonn).

Hassan, M., & Schneider, F. (2016). *Size and Development of the Shadow Economies of 157 Countries Worldwide: Updated and New Measures from 1999 to 2013* (IZA Discussion Paper No. 10281.).

Huynh, C. M., & Nguyen, T. L. (2019). Fiscal policy and shadow economy in Asian developing countries: does corruption matter? *Empirical Economics*, 1–17. https://doi.org/10.1007/s00181-019-01700-w

Ihrig, J., & Moe, K. S. (2004). Lurking in the shadows: the informal sector and government policy. *Journal of Development Economics*, *73*(2), 541–557. https://doi.org/10.1016/J.JDEVECO.2003.04.004

Johnson, S., Kaufmann, D., & Zoido-Lobaton, P. (1998). Regulatory Discretion and the Unofficial Economy. *American Economic Review*, *88*(2), 387–392.

Kirchgaessner, G. (2016). On estimating the size of the shadow economy". *German Economic Review*, *18*(1), 99-111.

La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2008). The economic consequences of legal origins. *Journal of Economic Literature*, *46*(2), 285–332. https://doi.org/10.1257/jel.46.2.285

La Porta, R., Lopez-De-Silanes, F., Shleifer, A., & Vishny, R. W. (1997). Legal Determinants of External Finance. *The Journal of Finance*, *52*(3), 1131–1150. https://doi.org/10.1111/j.1540-6261.1997.tb02727.x

La Porta, R., & Shleifer, A. (2014). Informality and Development. *Journal of Economic Perspectives*, *28*(3), 109–126. https://doi.org/10.1257/jep.28.3.109

Lewis, W. A. (1954). Economic Development with Unlimited Supplies of Labour. *The Manchester School*, *22*(2), 139–191. https://doi.org/10.1111/j.1467-9957.1954.tb00021.x

Loayza, N. (1999). *The Economics of the Informal Sector: A Simple Model and Some*

Empirical Evidence from Latin America. (World Bank Policy Research Working Paper No. 1727.).

Maloney, W. F. (2004). Informality Revisited. *World Development*, *32*(7), 1159–1178. https://doi.org/10.1016/J.WORLDDEV.2004.01.008

Massenot, B., & Straub, S. (2015). *Informal Sector and Economic Development: The Credit Supply Channel. SAFE Working Paper No. 106*. https://doi.org/10.2139/ssrn.2611570

Massenot, B., & Straub, S. (2016). Informal sector and economic development: The credit supply channel. *Economic Inquiry*, *54*(2), 1046–1067. https://doi.org/10.1111/ecin.12301

Medina, L., & Schneider, F. (2018). *Shadow Economies Around the World: What Did We Learn Over the Last 20 Years?* (IMF Working Paper No. 18/17).

Messaouda, B., & Teheni, Z. E. G. (2014). Business regulations and economic growth: What can be explained? *Business Regulations and Economic Growth: What Can Be Explained?*, *2*(2), 69–78.

Mishra, A., & Ray, R. (2010). *Informality, Corruption, and Inequality* (Department of Economics, University of Bath.).

Morrisson, C., & Murtin, F. (2011). *Internal income inequality and global inequality* (FERDI Working Paper.).

Perry, G., Maloney, W., Arias, O., Fajnzylber, P., Mason, A., & Saavedra-Chaduvi, J. (2007). *Informality, Exit and exclusion.* Washington D.C.: World Bank, Latin American and Caribbean Studies.

Phillips, P. C. B., & Sul, D. (2007). Transition Modeling and Econometric Convergence Tests. *Econometrica*, *75*(6), 1771–1855. https://doi.org/10.1111/j.1468-0262.2007.00811.x

Rauch, J. E. (1991). Modelling the informal sector formally. *Journal of Development*

Economics, *35*(1), 33–47. https://doi.org/10.1016/0304-3878(91)90065-4

Rosser, J. B., Rosser, M. V., & Ahmed, E. (2000). Income Inequality and the Informal

Economy in Transition Economies. *Journal of Comparative Economics*, *28*(1), 156–171.

https://doi.org/10.1006/JCEC.2000.1645

Rosser, J. B., Rosser, M. V., & Ahmed, E. (2003). Multiple unofficial economy equilibria

and income distribution dynamics in systemic transition. *Journal of Post Keynesian*

Economics, *25*(3), 425–447.

Sarte, P.-D. G. (2000). Informality and rent-seeking bureaucracies in a model of long-run

growth. *Journal of Monetary Economics*, *46*(1), 173–197.

https://doi.org/10.1016/S0304-3932(00)00020-9

Savoia, A., Easaw, J., & McKay, A. (2010). Inequality, democracy, and institutions: A

critical review of recent research. *World Development*, *38*(2), 142–154.

https://doi.org/10.1016/J.WORLDDEV.2009.10.009

Schneider, F. (2017). Estimating a Shadow Economy: Results, Methods, Problems, and Open

Questions. *De Gruyter Open, Open Economics*, *1*, 1–29.

Schneider, F., & Enste, D. H. (2013). *The shadow economy : An international survey.*

(Cambridge University Press., Ed.). Cambridge, UK.

Singer, H. W. (1970). Dualism revisited: A new approach to the problems of the dual society

in developing countries. *The Journal of Development Studies*, *7*(1), 60–75.

https://doi.org/10.1080/00220387008421348

Srinivas, H. (2016). *14 Reasons why the Informal Credit Market is used by the Poor: Policy*

Implications for Microcredit Programmes in Developing Countries. (Case Study Series

E-111).

Straub, S. (2005). Informal sector: The credit market channel. *Journal of Development*

Economics, *78*(2), 299–321. https://doi.org/10.1016/J.JDEVECO.2004.09.005

Tanzi, V. (1999). Uses and Abuses of Estimates of the Underground Economy. *The Economic Journal*, *109*(456), 338–347. https://doi.org/10.1111/1468-0297.00437

Tokman, V. (2007). *Modernizing the Informal Sector* (UN/DESA Working Paper No. 42).

Valentini, E. (2009). Underground Economy, Evasion and Inequality. *International Economic Journal*, *23*(2), 281–290. https://doi.org/10.1080/10168730902903433

Vanek, J., Chen, M. A., Carré, F., Heintz, J., & Hussmanns, R. (2014). *Statistics on the informal economy: Definitions, regional estimates & challanges. WIEGO working paper 2.*

World Bank. (2019). *Doing Business 2019. Training for Reform. Comparing Business Regulation for Domestic Firms in 190 Economies*. Washington, DC: United States: World Bank.

Wu, D. F., & Schneider, F. (2019). *Nonlinearity between the shadow economy and the level of development*.

Yap, W. W., Sarmidi, T., Shaari, A. H., & Said, F. F. (2018). Income inequality and shadow economy: a nonparametric and semiparametric analysis. *Journal of Economic Studies*, *45*(1), 2–13. https://doi.org/10.1108/JES-07-2016-0137

Appendix

Table A.1 Data sources

Variable	Description	Source
Gini index for years 1700, 1820, 1870, 1929, 1960, 1992	We interpolated the Gini index using the income deciles provided by Bourguignon and Morrisson (2002) and Morrisson and Murtin (2011) using the formula below, where k is the cumulated proportion of the population variable, for $k = 0,...,n$, with $X_0 = 0$, $X_n = 1$. Y_k is the cumulated proportion of the income variable, for $k = 0,...,n$, with $Y_0 = 0$, $Y_n = 1$. Y_k should be indexed in non-decreasing order ($Y_k > Y_{k-1}$) $$G = 1 - \sum_{k=1}^{n}(X_k - X_{k-1})(Y_k + Y_{k-1})$$	Own estimates using the decile's income share using the estimates by Bourguignon and Morrisson (2002) and Morrisson and Murtin (2011)
GDP per capita for year 1870	Gross Domestic Product per capita for year 1870	Maddison Project, Maddison (2010)
WIID Gini circa year 1980	Gini index for year circa 1980	WIID – World Income Inequality Database, UNU-WIDER
Informal economy 1999-2013	Size of the informal economy as percentage of GDP (%) using MIMIC method	Hassan and Schneider (2016)
Informal economy 1991–2015	Size of the informal economy as percentage of GDP (%) using MIMIC method and nights lights data to independently measure economic activity	Medina and Schneider (2018)
Employed in Informal Economy 2004–2016	Employed in the informal sector as a percent of employment (%)	ILOSTAT, 2018
Firms do not pay sales tax 2002–2009	Firms that do not report all sales for tax purposes are the percentage of firms that expressed that a typical firm reports less than 100 percent of sales for tax purposes.	Enterprise Surveys, World Bank
Lending interest rate 1995–2013	Lending rate is the bank rate that usually meets the short- and medium-term financing needs of	International Monetary Fund, International Financial Statistics and data files.
Getting credit score 2005–2013	Overall easiness to get credit 0-100. Index comprises strength of legal rights index and depth of credit information.	Doing Business 2018, World Bank
Cost of starting a business (% of income per capita) 200(	Cost of starting a business for men (% of income per capita)	Doing Business 2018, World Bank
Total tax rate (% of profit) 2006–2013	The total tax rate measures the amount of taxes and mandatory contributions payable by the business in the second year of operation, expressed as a share of commercial profits.	Doing Business 2018, World Bank
Origin of country's legal code		La Porta, Lopez-de-Silanes and Shleifer (2008)
List of countries and country groups used by Bourguignon and Morrisson (2002)	In Africa: Cote d'Ivoire-Ghana-Kenya, Egypt, Nigeria, North Africa (Algeria, Morocco, Tunisia, Libya, Mauritania, Sudan), South Africa, 46 African countries. In Asia: Bangladesh-Pakistan, Burma, China, India, Indonesia, Japan, Korea-Taiwan, Philippines-Thailand, and 45 Asian countries. In Latin America: Argentina-Chile, Brazil, Colombia-Peru-Venezuela, Mexico, and 37 Latin American countries. In Eastern Europe: Bulgaria-Greece-Romania-Yugoslavia-Croatia-Slovenia-Macedonia, Poland, Russia and Turkey. In Western Europe and European offshoots: Australia-Canada-New Zealand, Austria, Czechoslovakia, Austria-Hungary, France, Germany, Italy, Scandinavian countries (Denmark, Iceland, Sweden, Finland, Norway), Spain-Portugal, Benelux and microstates (Belgium, Malta, Netherlands, Luxemburg, Switzerland), United Kingdom-Ireland, and United States.	Bourguignon and Morrisson (2002)
Countries analysed as have data on both historic Gini indices and informal economy	In Africa: Algeria, Angola, Armenia, Benin, Botswana, Burkina Faso, Burundi, Cape Verde, Central Africa Rep., Chad, Comoros, Congo, Dem. R., Congo Rep. Of. Cote d'Ivorie, Equatorial Guinea, Ethiopia, Egypt, Gabon, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Libya, Madagascar, Malawi, Mali, Mauritania, Mauritius, Morocco, Mozambique, Nigeria, Rwanda, Senegal, Sierra Leone, South Africa, Swaziland, Tanzania, The Gambia, Togo, Tunisia, Uganda, Zambia and Zimbabwe. In Asia: Armenia, Azerbaijan, Bahrain, Bangladesh, Bhutan, Cambodia, Cameroon, China, Hong Kong SAR China, India, Indonesia, Iran, Japan, Jordan, Korea, Kuwait, Lao PDR, Lebanon, Macao, Malaysia, Mongolia, Myanmar, Nepal, Omar, Pakistan, Philippines, Qatar, Saudi Arabia, Singapore, Sri Lanka, Taiwan, Tajikistan, Thailand, Timor-Leste, Turkey, United Arab Em., Vietnam and Yemen. In Eastern Europe: Bulgaria, Croatia, Czech Republic, Hungary, Poland, Romania, Russia and Slovenia. In Western Europe: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Luxemburg, Malta, Netherlands, Norway, Portugal, Slovak Republic, Spain, Sweden, Switzerland and United Kingdom. In Latin America: Argentina, Belize, Bolivia, Brazil, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Guatemala, Haiti, Honduras, Mexico, Nicaragua, Paraguay, Peru, Suriname, Trinidad and Tobago, Uruguay and Venezuela. In North America: Canada and United States. In Oceania: Australia, Fiji and New Zealand.	Bourguignon and Morrisson (2002)

Table A.2 Summary statistics by region

	Africa			Asia			Eastern Europe			Western Europe			Latin America			North America			Oceania		
	Number of countries	Mean	SD	Number of countries	Mean	SD	Number of countries	Mean	SD	Number of countries	Mean	SD	Number of countries	Mean	SD	Number of countries	Mean	SD	Number of countries	Mean	SD
Gini index for year 1700	45	0.358	0.054	37	0.463	0.035	8	0.485	0.039	20	0.448	0.035	23	0.546	0.029	2	0.434	0.005	3	0.450	0.021
Gini index for year 1870	45	0.361	0.055	37	0.463	0.035	8	0.485	0.039	20	0.448	0.035	23	0.546	0.029	2	0.434	0.005	3	0.450	0.021
Gini index for year 1929	45	0.387	0.052	37	0.467	0.028	8	0.466	0.059	20	0.429	0.019	23	0.546	0.029	2	0.458	0.028	3	0.450	0.021
Gini index for year 1960	45	0.399	0.068	37	0.454	0.046	8	0.335	0.036	20	0.377	0.036	23	0.546	0.029	2	0.399	0.017	3	0.416	0.051
Gini index for year 1992	45	0.386	0.049	37	0.452	0.047	8	0.394	0.044	20	0.287	0.055	23	0.550	0.026	2	0.310	0.044	3	0.344	0.113
% Change in Gini indices for years 1700 and 1992	45	0.085	0.110	37	-0.024	0.073	8	-0.185	0.084	20	-0.361	0.108	23	0.008	0.026	2	-0.284	0.111	3	-0.242	0.210
Gini index for year circa 1980 (UNU-WIDER)	10	0.496	0.100	7	0.379	0.088	3	0.221	0.022	17	0.294	0.047	3	0.446	0.028	1	0.289	.	3	0.341	0.074
Informal economy 1999-2013 (Hassan and Schneider, 2016)	42	40.798	10.573	37	29.298	13.488	9	30.331	6.778	20	19.127	6.347	23	42.606	14.643	2	12.670	4.950	3	19.327	9.505
Informal economy 1991-2015 (Medina and Schneider, 2018)	45	39.571	7.764	37	28.545	11.749	8	26.974	6.672	20	16.149	6.380	23	38.089	10.901	2	11.130	3.946	3	18.743	11.889
Employed in Informal Economy 2004-2016	11	58.7773	20.6766	10	53.96	25.299	2	13.383	1.3906	0			16	57.834	12.91	0			0		
Firms do not pay sales tax 2002-2009	33	57.0935	21.5755	13	50.67	21.683	7	44.405	9.0688	5	34.901	12.809	15	47.009	14.144	0			0		
Lending interest rate 1999-2013	37	20.266	21.052	24	12.250	5.925	6	29.626	27.694	16	8.005	3.872	20	20.367	8.337	1	4.884	.	2	7.928	0.672
Getting credit score 2005-2013	44	33.844	16.355	37	49.533	22.170	8	67.109	17.002	20	64.313	19.745	23	53.228	18.430	2	87.500	8.839	3	83.542	13.583
Cost of starting a business (% of income per capita) 2004-2013	44	138.817	152.627	37	32.345	47.848	8	8.395	5.516	20	5.997	5.901	23	52.307	61.936	2	0.892	0.444	3	9.021	14.394
Total tax rate (% of profit) 2006-2013	44	65.555	58.649	37	36.017	16.010	8	40.475	10.578	20	43.701	13.658	23	49.341	20.479	2	38.249	7.834	3	40.701	7.233

Table A.3 First-stage regression of Table 4

Model using as covariate:	Gini 1700						Gini 1870						Gini 1929					
	Model 1			Model 2			Model 1			Model 2			Model 1			Model 2		
Dependent variable>	Getting credit score	Cost of starting a business	Total tax rate	Getting credit score	Cost of starting a business	Total tax rate	Getting credit score	Cost of starting a business	Total tax rate	Getting credit score	Cost of starting a business	Total tax rate	Getting credit score	Cost of starting a business	Total tax rate	Getting credit score	Cost of starting a business	Total tax rate
Excluded instruments:																		
Was not a colony	16.939***	-44.082	-17.945	19.320***	-50.676	-19.398	16.919**	-44.142	-18.080	19.289***	-50.341	-19.308	18.138***	-54.683	-21.378	19.940***	-55.791	-21.137
	(6.228)	(53.984)	(19.488)	(5.560)	(43.168)	(15.450)	(6.703)	(58.670)	(21.137)	(5.647)	(44.051)	(15.749)	(6.628)	(59.243)	(21.662)	(5.709)	(45.468)	(16.313)
British colony	2.437	-21.728	-17.109	3.700	-26.464	-17.678	2.465	-22.206	-17.263	3.679	-26.369	-17.646	3.039	-26.722	-18.782	4.071	-29.665	-18.786
	(7.399)	(65.494)	(25.354)	(5.831)	(51.006)	(19.702)	(7.995)	(71.127)	(27.503)	(5.941)	(52.072)	(20.093)	(7.879)	(70.467)	(27.449)	(5.994)	(53.080)	(20.561)
French colony	0.939	-39.935	-12.378	-0.661	-37.679	-12.399	0.684	-37.419	-11.325	-0.945	-34.814	-11.188	1.050	-39.278	-12.260	-0.582	-36.927	-12.088
	(7.838)	(65.030)	(24.327)	(6.114)	(50.038)	(18.818)	(8.400)	(69.418)	(26.002)	(6.176)	(50.169)	(18.909)	(8.179)	(69.133)	(25.778)	(6.176)	(51.400)	(19.260)
Spanish colony	18.988***	-17.101	-3.375	17.067***	-13.029	-3.072	18.638***	-15.134	-3.191	16.704***	-10.918	-2.978	18.777***	-18.946	-3.676	16.739**	-14.050	-3.498
	(6.648)	(50.076)	(15.825)	(5.324)	(38.513)	(12.418)	(7.152)	(54.248)	(17.038)	(5.411)	(39.288)	(12.660)	(6.914)	(52.296)	(16.631)	(5.406)	(38.948)	(12.731)
Portuguese colony	4.205	-13.886	-32.024	1.380	-6.841	-32.678	4.376	-14.566	-31.670	1.592	-8.004	-32.601	4.915	-17.125	-33.083	2.227	-11.523	-34.061
	(7.724)	(124.057)	(25.075)	(6.189)	(96.310)	(20.349)	(8.290)	(134.329)	(26.803)	(6.277)	(98.240)	(20.685)	(8.074)	(133.022)	(26.869)	(6.257)	(100.314)	(21.180)
Common law origin	9.814**	7.528	-11.294	21.050***	-22.084	-13.587	9.879**	6.989	-11.455	21.076***	-22.247	-13.591	10.325**	3.618	-12.634	21.450***	-25.120	-14.626
	(4.286)	(30.012)	(11.355)	(3.332)	(26.745)	(10.026)	(4.609)	(32.456)	(12.162)	(3.396)	(27.318)	(10.214)	(4.519)	(31.320)	(11.812)	(3.416)	(27.562)	(10.284)
Socialist legal origin	13.380*	22.132	2.519				-14.155**	28.274	4.347			4.827	-14.162**	27.532	4.348			
	(6.387)	(22.415)	(5.813)				(6.903)	(25.071)	(6.413)			(18.529)	(6.734)	(24.650)	(6.236)			
French legal orgin	-15.408	42.173*	3.195				-15.246	40.673	2.623				-15.008	38.595	1.988			
	(3.765)	(22.454)	(7.305)				(4.058)	(24.100)	(7.602)				(4.009)	(23.527)	(7.488)			
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number obs	1,117	1,117	1,117	1,117	1,117	1,117	1,117	1,117	1,117	1,117	1,117	1,117	1,117	1,117	1,117	1,117	1,117	1,117
Wald chi	642	129	86	784	143	93	614	126	83	780	143	91	617	126	83	792	141	89
Prob > chi2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table A.3 First-stage regression of Table 4 (Cont.)

Model using as covariate:	Gini 1960						Gini 1992					
	Model 1			Model 2			Model 1			Model 2		
Dependent variable>	Getting credit score	Cost of starting a business	Total tax rate	Getting credit score	Cost of starting a business	Total tax rate	Getting credit score	Cost of starting a business	Total tax rate	Getting credit score	Cost of starting a business	Total tax rate
Excluded instruments:												
Was not a colony	19.251***	-63.975	-24.606	20.550***	-68.357	-25.497	20.613***	-73.608	-28.044	19.427***	-51.593	-19.744
	(6.188)	(54.902)	(20.232)	(6.032)	(49.221)	(17.714)	(5.879)	(53.410)	(19.846)	(5.536)	(43.076)	(15.458)
British colony	2.288	-18.459	-16.071	3.593	-24.561	-16.996	2.954	-25.390	-18.380	3.828	-27.597	-18.074
	(7.061)	(62.394)	(24.041)	(6.166)	(54.267)	(20.816)	(6.634)	(58.583)	(22.922)	(5.817)	(50.848)	(19.690)
French colony	-1.581	-19.331	-5.515	-2.727	-18.811	-5.751	0.381	-32.727	-10.248	-0.659	-38.106	-12.457
	(7.147)	(56.573)	(21.557)	(6.224)	(48.788)	(18.600)	(6.915)	(56.680)	(21.423)	(6.100)	(49.948)	(18.783)
Spanish colony	21.734***	-22.266	-5.538	20.261***	-19.655	-5.433	21.247***	-41.002	-10.972	17.687***	-17.574	-4.896
	(6.312)	(49.331)	(15.811)	(5.618)	(42.395)	(13.692)	(5.855)	(46.358)	(15.234)	(5.283)	(38.194)	(12.531)
Portuguese colony	1.980	-4.288	-28.786	-0.802	4.503	-28.345	2.665	1.855	-27.504	1.274	-6.522	-32.452
	(7.478)	(117.203)	(22.449)	(6.699)	(101.301)	(19.594)	(6.909)	(109.334)	(21.081)	(6.218)	(95.977)	(19.944)
Common Law	8.898**	18.853	-7.601	20.523***	-15.831	-11.372	7.854**	22.050	-6.260	21.172***	-23.264	-13.980
	(4.245)	(31.982)	(11.713)	(3.588)	(29.687)	(10.988)	(4.003)	(31.083)	(11.565)	(3.309)	(26.511)	(9.954)
Socialist legal origin	-12.255	13.112	-0.452				-16.834***	45.789*	11.030			
	(6.084)	(20.227)	(5.209)				(6.084)	(27.469)	(7.876)			
French legal orgigin	-15.702***	49.850**	5.653				-16.699***	50.787**	6.334			
	(3.711)	(22.667)	(7.278)				(3.436)	(21.775)	(7.344)			
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number obs	1,117	1,117	1,117	1,117	1,117	1,117	1,117	1,117	1,117	1,117	1,117	1,117
Wald chi	602	132	84	684	140	85	627	135	90	782	144	92
Prob > chi2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

The top two rows show whether the first-stage regression corresponds to the Model 1 or Model 2 specifications, and whether using as main control the Gini index of the year 1700 or for any other year.

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table A.4 Historic inequality and Hassan and Schneider's informal economy, random effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Log Informal Economy							
LogGini index for year 1700	4.583***	2.791***	2.883***	2.729***	2.973***	3.028***	3.059***	2.973***
	(0.000)	(0.116)	(0.151)	(0.223)	(0.110)	(0.155)	(0.008)	(0.025)
% Change in Gini indices for years 1700 and 1992					1.260***	1.000***	1.489***	1.032***
					(0.070)	(0.026)	(0.042)	(0.007)
Lending interest rate		0.003		0.002	0.003			
		(0.004)		(0.004)	(0.004)			
Getting credit score			0.000	-0.001		0.000		
			(0.001)	(0.001)		(0.001)		
Cost of starting a business (% of income per capita)		0.000	-0.000	0.000	0.000	-0.000		
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)		
Total tax rate (% of profit)		-0.000	0.000	0.000	-0.000	0.000		
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)		
ΔLending interest rate $_{t-t-1}$							0.001	
							(0.002)	
ΔGetting credit score $_{t-t-1}$								0.001
								(0.001)
ΔCost of starting a business (% of income per capita) $_{t-t-1}$							0.000	-0.000
							(0.000)	(0.000)
ΔTotal tax rate (% of profit) $_{t-t-1}$							0.001	0.001
							(0.001)	(0.000)
Year fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,040	663	970	663	663	970	569	847
Number of countries	136	93	123	93	93	123	92	123

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.5 Alternative inequality indices and Hassan and Schneider's informal economy, random effects model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
										Log Informal Economy										
LogGini index for year 1700	2.973***	3.059***	3.028***	2.973***																
	(0.110)	(0.008)	(0.155)	(0.025)																
LogGini index for year 1870					2.973***	3.059***	3.028***	2.973***												
					(0.110)	(0.008)	(0.155)	(0.025)												
LogGini index for year 1929									2.948***	3.033***	3.002***	2.948***								
									(0.109)	(0.008)	(0.153)	(0.025)								
LogGini index for year 1960													2.403***	2.472***	2.447***	2.403***				
													(0.089)	(0.007)	(0.125)	(0.020)				
LogGini index for year 1992																	2.849***	2.931***	2.901***	2.848***
																	(0.106)	(0.008)	(0.148)	(0.024)
% Change in Gini indices for years 1700 and 1992	1.260***	1.489***	1.000***	1.032***	1.260***	1.489***	1.000***	1.032***	1.031***	1.253***	0.767***	0.803***	3.149***	3.433***	2.924***	2.921***	-2.952***	-2.845***	-3.289***	-3.179***
	(0.070)	(0.042)	(0.026)	(0.007)	(0.070)	(0.042)	(0.026)	(0.007)	(0.076)	(0.041)	(0.015)	(0.005)	(0.066)	(0.045)	(0.124)	(0.023)	(0.204)	(0.036)	(0.193)	(0.029)
Lending interest rate	0.003				0.003				0.003				0.003				0.003			
	(0.004)				(0.004)				(0.004)				(0.004)				(0.004)			
Getting credit score			0.000				0.000				0.000				0.000				0.000	
			(0.001)				(0.001)				(0.001)				(0.001)				(0.001)	
Cost of starting a business (% of income per capita)	0.000		-0.000		0.000		-0.000		0.000		-0.000		0.000		-0.000		0.000		-0.000	
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	
Total tax rate (% of profit)	-0.000		0.000		-0.000		0.000		-0.000		0.000		-0.000		0.000		-0.000		0.000	
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	
ΔLending interest rate $_{t-t-1}$		0.001				0.001				0.001				0.001				0.001		
		(0.002)				(0.002)				(0.002)				(0.002)				(0.002)		
ΔGetting credit score $_{t-t-1}$				0.001				0.001				0.001				0.001				0.001
				(0.001)				(0.001)				(0.001)				(0.001)				(0.001)
ΔCost of starting a business (% of income per capita)		0.000		-0.000		0.000		-0.000		0.000		-0.000		0.000		-0.000		0.000		-0.000
		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)
ΔTotal tax rate (% of profit) $_{t-t-1}$		0.001		0.001		0.001		0.001		0.001		0.001		0.001		0.001		0.001		0.001
		(0.001)		(0.000)		(0.001)		(0.000)		(0.001)		(0.000)		(0.001)		(0.000)		(0.001)		(0.000)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	663	569	970	847	663	569	970	847	663	569	970	847	663	569	970	847	663	569	970	847
Number of countries	93	92	123	123	93	92	123	123	93	92	123	123	93	92	123	123	93	92	123	123

Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.6 IV Estimates of inequality indices and Hassan and Schneider's informal economy, random effects model

	(1) Model 1	(2) Model 2	(3) Model 1	(4) Model 2	(5) Model 1	(6) Model 2	(7) Model 1	(8) Model 2	(9) Model 1	(10) Model 2
					Log Informal Economy					
LogGini index for year 1700	3.637***	2.512***								
	(0.036)	(0.000)								
LogGini index for year 1870			3.636***	2.512***						
			(0.037)	(0.000)						
LogGini index for year 1929					7.143**	-				
					(3.248)					
LogGini index for year 1960							1.416***	0.245***		
							(0.000)	(0.000)		
LogGini index for year 1992									2.427***	4.835***
									(0.018)	(0.034)
% Change in Gini indices for years 1700 and 1992		1.276***		1.276***		0.371***		0.624***		-
		(0.000)		(0.000)		(0.000)		(0.000)		
Getting credit score	-0.005***	-0.023***	-0.005***	-0.023***	0.018**	-0.032	-0.016***	-0.029***	-0.013***	0.007***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.009)	(.)	(0.000)	(0.000)	(0.000)	(0.000)
Cost of starting a business (% of income per capita)	-0.000***	-0.001***	-0.000***	-0.001***		-0.001	-0.001***	-0.001***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)		(.)	(0.000)	(0.000)	(0.000)	(0.000)
Total tax rate (% of profit)	0.007***	0.004***	0.007***	0.004***	0.013	0.002***	0.004***	0.002***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.013)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	939	909	939	909	939	970	939	970	970	939
Number of countries	119	115	119	115	119	123	119	123	123	119
Hausman test										
Chi2	101.170	269.870	280.240	269.870	28.750	24.720	6.690	51.250	0.000	55.210
Prob>chi2	0.915	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.999	1.000

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.7 First-stage regression of Table A.6

Model using as covariate:	Gini 1700						Gini 1870						Gini 1929				
	Model 1			Model 2			Model 1			Model 2			Model 1		Model 2		
Dependent variable>	Getting credit score	Cost of starting a business	Total tax rate	Getting credit score	Cost of starting a business	Total tax rate	Getting credit score	Cost of starting a business	Total tax rate	Getting credit score	Cost of starting a business	Total tax rate	Getting credit score	Total tax rate	Getting credit score	Cost of starting a business	Total tax rate
Excluded instruments:																	
Was not a colony	110.879***	-121.434***	-45.1702**				-13.778**	97.731	-17.56984				7.031	2.864	4.298**	-1.361	1.655
	(7.760)	(19.663)	(18.240)				(6.741)	(17.793)	(12.588)				(4.3E+08)	(1.3E+08)	(1.960)	(11.366)	(1.409)
Dutch colony				14.362***	-44.580**	7.681852				-45.744***	42.230**	46.788***	-112.588	-71.0549			
				(5.606)	(21.497)	(19.260)				(5.325)	(15.012)	(1.972)	(8.8E+08)	(2.5E+08)			
Spanish colony	1.016263	-45.027***	-103.502***	16.613***	-75.659***	11.58407	38.133***	-58.949***	18.631***	-19.913***	-22.906**	35.348***	-50.960	-37.536	22.981***	119.152***	7.054***
	(5.027)	(17.066)	(18.634)	(4.502)	(15.661)	(11.772)	(6.564)	(11.625)	(6.536)	(3.028)	(10.540)	(1.645)	(8.E+08)	(2.1E+08)	(2.121)	(38.240)	(1.265)
British colony				13.635**	144.313***	22.63774											
				(5.153)	(20.879)	(17.068)											
French colony	43.136***	-89.822***	-54.807**	17.686**	-83.442***	15.34675	-40.137***	77.524***	-17.769***	-17.659***	-32.393***	38.343***	-16.641	-2.907	57.299***	166.030***	41.684***
	(7.035)	(16.183)	(18.648)	(7.293)	(15.384)	(11.466)	(5.246)	(12.840)	(6.500)	(4.738)	(9.707)	(1.734)	(4.7E+08)	(1.2E+08)	(7.715)	(40.411)	(2.082)
Portuguese colony	-34.375***	127.115***	22.65749	1.807831	-34.896**	38.484***	-34.375***	127.115***	22.65749	-27.866***	-14.06661	58.977***	-50.960	133.264	-34.375***	127.115***	22.65749
	(4.948)	(18.887)	(18.633)	(4.690)	(11.471)	(2.999)				(3.960)	(11.223)	(1.200)	(8.E+08)	(2.1E+08)	(4.948)	(18.694)	(18.638)
German colony	-34.375***	26.106**	-13.23478	1.807831	135.906***	2.592001	-34.375***	26.106**	-13.235***	-49.004***	-62.519***	35.652***	-55.647	-47.622	-34.375***	26.106**	-13.235**
	(4.948)	(12.534)	(5.857)	(5.097)	(21.654)	(17.261)	(4.941)	(18.789)	(18.620)	(4.335)	(15.915)	(5.911)	(7.9E+08)	(2.1E+08)	(4.948)	(12.669)	(5.864)
Belgium colony	-58.124***	93.584***	-55.496***	-41.292***	76.900***	166.994***	-58.124***	93.584***	-55.497***	-54.237***	95.596***	175.417***	-50.342	-41.4651			
	(9.586)	(25.260)	(2.489)	(4.750)	(22.836)	(37.455)	(4.941)	(12.546)	(5.838)	(3.772)	(21.840)	(37.218)	(7.5E+08)	(1.9E+08)			
Russia colony	-43.279***	26.79847	-224.689***	-26.449***	10.11522	-2.19788	-89.437***	180.222***	-65.519***	-24.601***	689.383***	292.973***	-34.645	-21.702			
	(12.266)	(39.867)	(37.238)	(8.833)	(25.809)	(5.084)	(9.584)	(25.624)	(2.531)	(3.291)	(96.217)	(11.677)	(1.E+09)	(2.2E+08)			
Common law origin				36.183***	162.011***	15.82678				-14.629**	-88.625***	48.886***					
				(6.768)	(20.117)	(16.294)	(13.244)	(36.053)	(12.955)	(6.215)	(13.751)	(1.574)					
French legal origen	53.671***	-80.811***	110.998***	11.906***	-22.109*	-8.486**				5.547***	-12.92443	-4.348749	-6.940	-26.022	31.707***	-6.685	0.441
	(5.027)	(17.066)	(18.634)	(2.546)	(12.337)	(4.311)				(2.372)	(12.143)	(3.803)	(1.8E+08)	(9.E+08)	(2.119)	(38.241)	(1.264)
Socialism legal origin				-40.530***	72.651***	-12.079**	-16.555***	66.888***	11.134***	-23.309***	47.777***	-23.285***					
				(3.161)	(12.646)	(6.067)	(4.918)	(9.955)	(6.379)	(4.024)	(11.788)	(2.610)					
German legal origin				-6.976*	-60.606***	-24.300***	-17.629***	-22.909**	-12.718***	-33.900***	-21.719**	-6.783***					
				(4.039)	(13.668)	(8.797)	(4.412)	(12.922)	(6.860)	(2.926)	(11.698)	(1.881)					
Scandinavia legal origin				-46.424***	-6.6863	-25.393***	15.965***	-22.402**	7.902463	-48.770***	-3.298638	-23.867***					
				(2.261)	(10.064)	(1.515)	(2.940)	(10.591)	(6.766)	(2.192)	(9.984)	(1.388)					
							-30.134***	-4.630048	-12.129***								
							(4.488)	(14.992)	(1.750)								
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes	Yes	Yes
Number obs	939	939	939	939	939	939	939	939	939	939	939	939	939	939	939	939	939
Wald chi	34357	10722	106866	34203	9315	81600	34203	9315	81600	34203	9315	81600	306	29	34133	9301	108954
Prob > chi2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.090	0.513

Table A.7 First-stage regression of Table A.6 (Cont.)

Model using as covariate:	Gini 1960						Gini 1992					
	Model 1			Model 2			Model 1			Model 2		
	Cost of			Cost of			Cost of			Cost of		
	Getting credit score	starting a business	Total tax rate	Getting credit score	starting a business	Total tax rate	Getting credit score	starting a business	Total tax rate	Getting credit score	starting a business	Total tax rate
Dependent variable>												
Excluded instruments: Was not a colony	15.269***	16.580*	14.31637	26.735***	4.630965	42.110***	-1.563	22.982	33.20***	19.332**	-39.282*	-144.262***
	(5.096)	(9.876)	(1.026)	(5.928)	(18.342)	(2.060)	(4.399)	(15.498)	(1.792)	(8.990)	(23.368)	(18.299)
Dutch colony												
Spanish colony	62.174***	-135.295***	-59.20555	78.889***	-62.0306	153.0639	56.256***	-44.944	104.788***	48.035***	-139.070***	
	(4.103)	(15.295)	(18.613)	(18.830)	(49.182)	(5.497)	(12.560)	(32.222)	(3.44)	(5.544)	(17.852)	
British colony												-84.698***
												(18.640)
French colony	-16.095***	0.96033	-95.631	14.851**	-30.0861	17.672***	37.374***	-15.349	59.365***	-30.233***	-2.815	
	(5.972)	(14.700)	(18.645)	(4.853)	(12.646)	(1.634)	(8.067)	(20.605)	(2.268)	(7.040)	(17.342)	
Portuguese colony	-34.375***	127.115***	22.65749	86.702***	-66.7653	163.897***				-34.375***	127.115***	121.1235***
	(4.948)	(18.887)	(18.633)	(19.053)	(49.736)	(5.685)				(4.948)	(18.887)	(18.673)
German colony	-34.375***	26.106**	-13.23478	-18.852***	-15.5845	-17.479***				-34.375***	26.106**	22.657**
	(4.948)	(12.534)	(5.857)	(4.922)	(12.025)	(1.929)				(4.948)	(12.534)	(18.633)
Belgium colony	-43.093***	97.598***	-28.39368							-55.939***	94.167***	-13.235**
	(7.058)	(19.531)	(1.763)							(9.185)	(24.335)	(5.857)
Russia colony	-28.249***	30.813	-197.5863							-41.096***	27.38179	-51.558***
	(10.410)	(36.507)	(37.197)							(11.955)	(39.287)	(2.375)
Common law origin												-220.751***
												(37.231)
French legal origen	-7.486**	9.457	66.7014	-2.352	-63.917***	-57.671***	-1.569	-80.894**	-97.292***	6.652118	13.23293	92.194***
	(3.675)	(13.089)	(18.573)	(8.434)	(20.705)	(2.416)	(12.560)	(32.222)	(3.440)	(5.235)	(16.003)	(18.601)
Socialism legal origin				21.850*	-0.1102	87.897***						
				(11.508)	(30.645)	(3.475)						
German legal origin				-33.594***	-0.50746	-20.621***						
				(4.236)	(10.116)	(1.766)						
Scandinavia legal origin												
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number obs	939	939	939	939	939	939	939	939	939	939	939	939
Wald chi	34357	10722	106866	34133	9301	108954	34133	9301	108954	34357	10722	106866
Prob > chi2	0.000	0.000	0.000	0.000	0.090	0.512	0.000	0.000	0.000	0.000	0.000	0.000

The top two rows show whether the first-stage regression corresponds to the Model 1 or Model 2 specifications, and whether using as main control the Gini index of the year 1700 or for any other year.

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.8 IV estimates of inequality and labour force employed in the informal economy as estimated by the International Labour Organisation (ILO) using labour surveys

	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 1	(5) Model 2	(6) Model 3	(7) Model 1	(8) Model 2	(9) Model 3	(10) Model 1	(11) Model 2	(12) Model 3	(13) Model 1	(14) Model 2	(15) Model 3
						Log Employed in Informal Economy									
LogGini index for year 1700	4.427***	5.219***	0.753***												
	(1.110)	(1.579)	(0.084)												
LogGini index for year 1870				4.427***	5.219***	0.753***									
				(1.110)	(1.579)	(0.084)									
LogGini index for year 1929							6.797***	5.479***	0.753***						
							(1.587)	(0.311)	(0.084)						
LogGini index for year 1960										6.797***	5.479***	0.642***			
										(1.587)	(0.311)	(0.071)			
LogGini index for year 1992													3.238***	9.031***	0.752***
													(0.634)	(0.568)	(0.084)
% Change in Gini indices for years 1700 and 1992		-	16.532***		-	16.532***		-	16.532***		-	16.182***		-	15.783***
			(4.728)			(4.728)			(4.728)			(4.720)			(4.711)
Getting credit score	-0.064***	-0.072***		-0.064***	-0.072***		-0.059***	0.077***		-0.059***	0.077***		-0.045***	0.123***	
	(0.015)	(0.018)		(0.015)	(0.018)		(0.015)	(0.007)		(0.015)	(0.007)		(0.008)	(0.007)	
Cost of starting a business (% of income per capita)	0.003	0.005		0.003	0.005		0.018***	0.015***		0.018***	0.015***		0.002	0.022***	
	(0.005)	(0.005)		(0.005)	(0.005)		(0.004)	(0.001)		(0.004)	(0.001)		(0.003)	(0.001)	
Total tax rate (% of profit)	-0.089***	-0.074***		-0.089***	-0.074***		0.189***	0.046***		0.189***	0.046***		-0.059***	0.034***	
	(0.024)	(0.009)		(0.024)	(0.009)		(0.054)	(0.005)		(0.054)	(0.005)		(0.013)	(0.004)	
ΔLending interest rate $_{t,t-1}$			0.044			0.044			0.044			0.044			0.044
			(0.062)			(0.062)			(0.062)			(0.062)			(0.062)
ΔCost of starting a business (% of income per capita)			-0.006			-0.006			-0.006			-0.006			-0.006
			(0.013)			(0.013)			(0.013)			(0.013)			(0.013)
ΔTotal tax rate (% of profit) $_{t,t-1}$			-0.002			-0.002			-0.002			-0.002			-0.002
			(0.008)			(0.008)			(0.008)			(0.008)			(0.008)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	90	90	35	90	90	35	90	90	35	90	90	35	90	90	35
Number of countries	30	30	8	30	30	8	30	30	8	30	30	8	30	30	8

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.9 First-stage regression of Table A.8

Dependent variable>	Gini 1700									Gini 1870								
	Model 1			Model 2			Model 3			Model 1			Model 2			Model 3		
	Getting credit score	Cost of starting a business	Total tax rate	Getting credit score	Cost of starting a business	Total tax rate	ΔLending interest rate $_{t-t-1}$	ΔCost of starting a business $_{t-t-1}$	ΔTotal tax rate $_{t-t-1}$	Getting credit score	Cost of starting a business	Total tax rate	Getting credit score	Cost of starting a business	Total tax rate	ΔLending interest rate $_{t-t-1}$	ΔCost of starting a business $_{t-t-1}$	ΔTotal tax rate $_{t-t-1}$
Excluded instruments:																		
Was not a colony	-6.25	0.037***	23.948***	-6.25	0.037***	23.948***				-6.25	0.0370***	23.948***	-6.25	0.037***	23.948***			
	(.)	(0.000)	(0.000)	(.)	(0.000)	(0.000)				(.)	(0.000)	(0.000)	(.)	(0.000)	(0.000)			
British colony	-21.451***	270.997***	17.97*	-21.451***	270.997***	17.97*				-21.451***	270.997***	17.97*	-21.451***	270.997***	17.970*			
	(2.574)	(12.566)	(8.606)	2.574	(12.566)	(8.606)				(2.574)	(12.566)	(8.606)	(2.574)	(12.566)	(8.606)			
Spanish colony	15.294***	258.839***	14.504	15.294***	258.839***	14.504				15.294***	258.839***	14.504	15.294***	258.839***	14.504			
	(2.591)	(12.920)	(8.834)	2.591	(12.920)	(8.834)				(2.591)	(12.920)	(8.834)	(2.591)	(12.920)	(8.834)			
Common law origin	-0.23	5.835***	-15.508***	-0.23	5.835***	-15.508***				-0.23	5.835***	-15.508***	-0.230	5.835***	-15.508***			
	(0.317)	(1.171)	(3.464)	0.317	(1.171)	(3.464)				(0.317)	(1.171)	(3.464)	(0.317)	(1.171)	(3.464)			
French legal orgin	-27.575***	-241.574***	-0.935	-27.575***	-241.574***	-0.935				-27.575*	-241.574***	-0.935	-27.575***	-241.574***	-0.935			
	(2.564)	(12.888)	(9.150)	2.564	(12.888)	(9.150)				(2.564)	(12.888)	(9.150)	(2.564)	(12.888)	(9.150)			
Lending interest rate in year 1985							0.017	-0.078*	0.085							0.017	-0.078	0.085
							(1.1E+5)	(2.2E+5)	(1.2E+6)							(1.1E+5)	(2.2E+5)	(1.2E+5)
Lagged $_{t-1}$ Cost of starting a business							0.042	-0.609**	-0.04							0.042	-0.609*	-0.04
							(0.158)	(0.320)	(0.133)							(0.158)	(0.320)	(0.133)
Lagged $_{t-1}$ total tax rate							0.116*	0.205	-0.857***							0.116*	0.205	-0.857***
							(0.070)	(0.151)	(0.098)							(0.070)	(0.151)	(0.098)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number observations	90	90	90	90	90	90	35	35	35	90	90	90	90	90	90	35	35	35
Wald chi	173145	1.00E+09	2.00E+15	173145	1.00E+09	2.00E+15	24740	84	154	173145	1.00E+09	2.00E+15	173145	1.00E+09	2.00E+15	24740	84	154
Prob > chi2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table A.9 First-stage regression of Table A.8 (Cont.)

	Gini 1929									Gini 1960								
	Model 1			Model 2			Model 3			Model 1			Model 2			Model 3		
Dependent variable>	Getting credit score	Cost of starting a business	Total tax rate	Getting credit score	Cost of starting a business	Total tax rate	ΔLending interest rate $_{t-t-1}$	ΔCost of starting a business $_{t-t-1}$	ΔTotal tax rate $_{t-t-1}$	Getting credit score	Cost of starting a business	Total tax rate	Getting credit score	Cost of starting a business	Total tax rate	ΔLending interest rate $_{t-t-1}$	ΔCost of starting a business $_{t-t-1}$	ΔTotal tax rate $_{t-t-1}$
Excluded instruments:																		
Was not a colony	-6.25	0.037***	23.948***	-6.25	0.037***	23.948***				-6.25	0.037***	23.948***	-6.25	0.037***	23.948***			
	(.)	(0.000)	(0.000)	(.)	(0.000)	(0.000)				(.)	(0.000)	(0.000)	(.)	(0.000)	(0.000)			
British colony	-21.451***	270.997***	17.97*	-21.451***	270.997***	17.97*				-21.451***	270.997***	17.97*	-21.451***	270.997***	17.97*			
	(2.574)	(12.566)	(8.606)	(2.574)	(12.566)	(8.606)				(2.574)	(12.566)	(8.606)	(2.574)	(12.566)	(8.606)			
Spanish colony	15.294***	258.839***	14.504	15.294***	258.839***	14.504				15.294***	258.839***	14.504	15.294***	258.839***	14.504			
	(2.591)	(12.920)	(8.834)	(2.591)	(12.920)	(8.834)				(2.591)	(12.920)	(8.834)	(2.591)	(12.920)	(8.834)			
Common law origin	-0.23	5.835***	-15.508***	-0.23	5.835***	-15.508***				-0.23	5.835***	-15.508***	-0.23	5.835***	-15.508***			
	(0.317)	(1.171)	(3.464)	(0.317)	(1.171)	(3.464)				(0.317)	(1.171)	(3.464)	(0.317)	(1.171)	(3.464)			
French legal orgin	-27.575***	-241.574***	-0.935	-27.575***	-241.574***	-0.935				-27.575***	-241.574***	-0.935	-27.575***	-241.574***	-0.935			
	(2.564)	(12.888)	(9.150)	(2.564)	(12.888)	(9.150)				(2.564)	(12.888)	(9.15)	(2.564)	(12.888)	(9.150)			
Lending interest rate in year 1985							0.017	-0.078	0.085							0.024	-0.109	0.05
							(1.1E+5)	(2.2E+5)	(1.2E+5)							(1.3E+5)	(2.3E+5)	(1.3E+5)
Lagged $_{t-1}$ Cost of starting a business							0.042	-0.609	-0.04							0.042	-0.609*	-0.04
							(0.158)	(0.320)	(0.133)							(0.158)	(0.320)	(0.133)
Lagged $_{t-1}$ total tax rate							0.116*	0.205	-0.857***							0.116*	0.205	-0.857*
							(0.070)	(0.151)	(0.098)							(0.070)	(0.151)	(0.098)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number observations	90	90	90	90	90	90	35	35	35	90	90	90	90	90	90	35	35	35
Wald chi	173145	1.00E+09	2.00E+15	173145	1.00E+09	2.00E+15	24740	84	154	173145	1.00E+09	2.00E+15	173145	1.00E+09	2.00E+15	24740	84	154
Prob > chi2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table A.9 First-stage regression of Table A.8 (Cont.)

Dependent variable>	Gini 1992								
	Model 1			Model 2			Model 3		
	Getting credit score	Cost of starting a business	Total tax rate	Getting credit score	Cost of starting a business	Total tax rate	ΔLending interest rate $_{t-t-1}$	ΔCost of starting a business $_{t-t-1}$	ΔTotal tax rate $_{t-t-1}$
Excluded instruments:									
Was not a colony	-6.25	0.037***	23.948***	-6.25	0.037***	23.948***			
	(.)	(0.000)	(0.000)	(.)	(0.000)	(0.000)			
British colony	-21.451***	270.997***	17.97*	-21.451***	270.997***	17.97*			
	(2.574)	(12.566)	(8.606)	(2.574)	(12.566)	(8.606)			
Spanish colony	15.294***	258.839***	14.504	15.294***	258.839***	14.504			
	(2.591)	(12.920)	(8.834)	(2.591)	(12.920)	(8.834)			
Common law origin	-0.230	5.835***	-15.508***	-0.23	5.835***	-15.508***			
	(0.317)	(1.171)	(3.464)	(0.317)	(1.171)	(3.464)			
French legal orgin	-27.575***	-241.574***	-0.935	-27.575***	-241.574***	-0.935			
	(2.564)	(12.888)	(9.150)	(2.564)	(12.888)	(9.150)			
Lending interest rate in year 1985							0.032	-0.145	0.011
							(1.5E+5)	(2.6E+5)	(1.5E+5)
Lagged $_{t-1}$ Cost of starting a business							0.042	-0.609*	-0.04
							(0.158)	(0.320)	(0.133)
Lagged $_{t-1}$ total tax rate							0.116*	0.205	-0.857***
							(0.070)	(0.151)	(0.098)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number observations	90	90	90	90	90	90	35	35	35
Wald chi	173145	1.00E+09	2.00E+15	173145	1.00E+09	2.00E+15	24740	84	154
Prob > chi2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

The top two rows show whether the first-stage regression corresponds to the Model 1, Model 2 or Model 3 specifications, and whether using as main control the Gini index of the year 1700 or for any other year.

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.10 IV estimates of inequality and firms that do not pay sales taxes as estimated by the World Bank based on business surveys

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Log Firms do not Pay Sales Tax									
LogGini index for year 1700	1.321***	4.156***								
	(0.000)	(1.294)								
LogGini index for year 1870			1.321***	4.156***						
			(0.000)	(1.294)						
LogGini index for year 1929					3.907***	1.509***				
					(1.066)	(0.000)				
LogGini index for year 1960							0.945***	0.801***		
							(0.000)	(0.000)		
LogGini index for year 1992									1.813***	3.778***
									(0.000)	(0.046)
% Change in Gini indices for years 1700 and 1992		-0.327		-0.327		-0.237***		3.674***		-4.402***
		(0.333)		(0.333)		(0.000)		(0.000)		(0.199)
Getting credit score	0.017***	-0.002	0.017***	-0.002	-0.002	0.016***	0.000***		0.003***	
	(0.000)	(0.008)	(0.000)	(0.008)	(0.008)	(0.000)	(0.000)		(0.000)	
Cost of starting a business (% of income per capita)	0.001***	-0.000	0.001***	-0.000	-0.000	0.001***	0.001***	0.001***	0.001***	0.000
	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Total tax rate (% of profit)	0.003***		0.003***			0.003***	0.002***	-0.003***	0.001***	
	(0.000)		(0.000)			(0.000)	(0.000)	(0.000)	(0.000)	
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	299	299	299	299	299	299	299	299	299	299
Number of countries	75	75	75	75	75	75	75	75	75	75
Hausman test										
Chi2	282.630	55.690	282.630	1811.500	409.420	36.890	49.080	1092.000	3715.650	425.930
Prob>chi2	0.000	0.909	0.000	0.000	0.000	1.000	0.989	0.000	0.000	1.000

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.11 First-stage regression of Table A.10

	Gini 1700						Gini 1870					
	Model 1			Model 2			Model 1			Model 2		
Dependent variable>	Getting credit score	Cost of starting a business	Total tax rate	Getting credit score	Cost of starting a business	Total tax rate	Getting credit score	Cost of starting a business	Total tax rate	Getting credit score	Cost of starting a business	Total tax rate
Excluded instruments:												
Was not a colony	6.879***	-148.641***	-16.69***	-17.776	41.646	-8.12	6.879***	-148.641***	-16.690***	-17.776	41.646	-8.12
	(1.217)	(51.506)	(0.483)	(1.9E+6)	(1.3E+6)	(1.3E+6)	(1.217)	(51.506)	(0.483)	(1.9E+6)	(1.3E+6)	(1.3E+6)
British colony	-3.043*	-68.449***	5.3***	-14.908	60.625	5.944	-3.043*	-68.449***	5.300***	-14.908	60.625	5.944
	(1.747)	(26.749)	(1.980)	(2.4E+6)	(1.8E+6)	(2.7E+6)	(1.747)	(26.749)	(1.980)	(2.4E+6)	(1.8E+6)	(2.7E+6)
Spanish colony	29.688***	-57.154***	-2.483***	28.991	18.47	-8.76	29.688***	-57.154***	-2.483*	28.991	18.47	-8.76
	(1.712)	(9.493)	(1.320)	(2.4E+6)	(1.8E+6)	(2.0E+6)	(1.712)	(9.493)	(1.320)	(2.4E+6)	(1.8E+6)	(2.0E+6)
Portuguese colony	13.129***	-154.676***	7.298***	-11.526	35.611	15.868	13.129***	-154.676***	7.298***	-11.526	35.611	15.868
	(1.777)	(52.227)	(0.681)	(2.4E+6)	(1.7E+6)	(1.4E+6)	(1.777)	(52.227)	(0.681)	(2.4E+6)	(1.7E+6)	(1.4E+6)
Common law origin	5.732***	194.461***	-3.041***	40.882	-46.055	-18.115	5.732***	194.461***	-3.041***	40.882	-46.055	-18.115
	(1.165)	(73.653)	(0.480)	(1.9E+6)	(1.5E+6)	(1.7E+6)	(1.165)	(73.653)	(0.480)	(1.9E+6)	(1.5E+6)	(1.7E+6)
Socialist law origin	-5.317*	153.951***	16.631***	-14.908	4.725	34.005	-5.317*	153.951***	16.631***	-14.908	4.725	34.005
	(3.105)	(52.927)	(0.604)	(2.3E+6)	(1.6E+6)	(1.1E+6)	(3.105)	(52.927)	(0.604)	(2.3E+6)	(1.6E+6)	(1.1E+6)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number observations	299	299	299	299	299	299	299	299	299	299	299	299
Wald chi	18534	6668	283942	53	31	21	18534	6668	283942	53	31	21
Prob > chi2	0.00	0.00	0.00	0.98	1.00	1.00	0.00	0.00	0.00	0.98	1.00	1.00

Table A.11 First-stage regression of Table A.10 (Cont.)

	Gini 1929						Gini 1960						Gini 1992				
	Model 1			Model 2			Model 1			Model 2			Model 1				
Dependent variable>	Getting credit score	Cost of starting a business	Total tax rate	Getting credit score	Cost of starting a business	Total tax rate	Getting credit score	Cost of starting a business	Total tax rate	Getting credit score	Cost of starting a business	Total tax rate	Getting credit score	Cost of starting a business	Total tax rate	Cost of starting a business	Total tax rate
Excluded instruments:																	
Was not a colony	-21.202	73.355	25.501	-39.024***	142.956***	-59.911***	4.234	5.221	-43.334***	-174.741**	20.616***		69.123***	-343.701***	-7.008***	15.683***	2.529
	(5.1E+6)	(2.6E+6)	(1.3E+6)	(4.263)	(45.287)	(1.127)	(5.506)	(22.305)	(1.438)	(69.091)	(1.722)		(2.108)	(91.966)	(1.533)	(7.260)	(5.4E+5)
British colony	-29.964	43.639	-34.975	-4.567**	7.419	3.407*	8.347*	-143.862**	23.325***	-143.862**	23.325***		25.779***	-210.730***	0.286	92.340***	-5.259
	(4.8E+6)	(2.9E+6)	(3.7E+6)	(1.791)	(19.908)	(1.892)	(4.748)	(63.237)	(1.393)	(63.237)	(1.393)		(1.233)	(66.301)	(1.003)	(1.780)	(2.1E+6)
Spanish colony	-34.55	114.42	47.444	5.140**	141.18***	-2.908*							29.688***	-57.154***	-2.483*		
	(6.2E+6)	(3.6E+6)	(3.0E+6)	(2.125)	(37.488)	(1.763)							(1.712)	(9.493)	(1.320)		
Portuguese colony	3.456	-39.042	-122.44				-16.574***	-137.344**	-11.733***				25.498***	-199.988**	-10.105***		
	(4.0E+6)	(2.5E+6)	(1.9E+6)				(4.963)	(63.656)	(1.502)				(1.916)	(66.183)	(1.235)		
Common law origin	1.894	96.581	120.733	22.168***	185.915***	-5.131***	-11.472**	175.773***	-5.984***	175.773***	-5.984***		1.194	172.981**	17.502***	-107.804***	2.856
	(3.0E+6)	(2.6E+6)	(1.7E+6)	(1.266)	(69.847)	(1.044)	(4.719)	(63.233)	(1.397)	(63.233)	(1.397)		(2.894)	(60.005)	(1.149)	(1.960)	(2.3E+6)
Socialist law origin	-39.419	13.692	14.873	-10.687***	-7.833	32.549***	27.880***	-197.785***	70.665***				-60.922***	201.308***	1.863		
	(4.3E+6)	(2.8E+6)	(7.4E+6)	(1.996)	(14.603)	(0.967)	(6.636)	(65.684)	(1.524)				(5.238)	(53.898)	(2.810)		
French legal orgin													-29.407***	59.853***	21.858***		
													(3.002)	(15.544)	(2.013)		
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number observations	299	299	299	299	299	299	299	299	299	299	299	299	299	299	299	299	299
Wald chi	53	31	23	18534	6668	283942	18534	6668	283942	6668	283942		18534	6668	283942	23	23
Prob > chi2	0.9813	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		0.0000	0.0000	0.0000	1.0000	1.0000

The top two rows show whether the first-stage regression corresponds to the Model 1 or Model 2 specifications, and whether using as main control the Gini index of the year 1700 or for any other year.

Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.12 Inequality and Hassan and Schneider's informal economy only for countries with available WIID Gini index, random effects model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
								Log Average Informal Economy									
Log WIID Gini index circa year 1980	1.033***	0.693*	5.568***	5.770***		0.103	5.807										
	(0.162)	(0.355)	(0.000)	(0.725)		(1.547)	(4.292)										
Squared log WIID Gini index circa year 1980						-0.283	2.176										
						(0.711)	(1.706)										
% Change in WIID Gini and indice for years 1980 and historical Gini 1700				-16.042***	-16.345***												
				(0.000)	(1.902)												
LogGini index for year 1700								1.949***	2.163***		4.974	18.398					
								(0.000)	(0.349)		(5.470)	(16.697)					
Squared log Gini index year 1700											2.905	11.586					
											(3.550)	(10.709)					
LogGini index for year 1992													1.950***	2.164***		0.105	0.272
													(0.000)	(0.349)		(1.902)	(1.595)
Squared log Gini index year 1992																-0.178	0.049
																(0.915)	(0.737)
% Change in Gini indices for years 1992 and 1700								-4.887***	-4.850***				-8.327***	-8.667***			
								(0.000)	(0.493)				(0.000)	(1.108)			
Getting credit score		-0.000		-0.000	-0.000	-0.000	0.001		-0.000	-0.000	-0.001	0.001		-0.000	-0.000	-0.000	0.001
		(0.001)		(0.001)	(0.001)	(0.001)	(0.002)		(0.001)	(0.001)	(0.001)	(0.002)		(0.001)	(0.001)	(0.001)	(0.002)
Cost of starting a business (% of income per capita)		-0.000*		-0.000	-0.000*	-0.000*	0.003		-0.000	-0.000	-0.000*	0.003		-0.000	-0.000	-0.000*	0.003
		(0.000)		(0.000)	(0.000)	(0.000)	(0.004)		(0.000)	(0.000)	(0.000)	(0.004)		(0.000)	(0.000)	(0.000)	(0.004)
Total tax rate (% of profit)		0.000		-0.000	0.000	0.000	0.001		-0.000	-0.000	0.000	0.001		-0.000	-0.000	0.000	0.001
		(0.000)		(0.000)	(0.000)	(0.000)	(0.002)		(0.000)	(0.000)	(0.000)	(0.002)		(0.000)	(0.000)	(0.000)	(0.002)
Log GDP per capita for year 1870							-0.562***					-0.713***					-0.618***
							(0.197)					(0.204)					(0.181)
Africa x Log WIID Gini index circa year 1980					-0.090												
					(0.264)												
Asia x Log WIID Gini index circa year 1980					7.022***												
					(1.366)												
Western Europe x Log WIID Gini index circa year 1980					0.964*												
					(0.521)												
Latin America x Log WIID Gini index circa year 1980					5.334***												
					(0.202)												
North America x Log WIID Gini index circa year 1980					0.000												
					(0.000)												
Oceania x Log WIID Gini index circa year 1980					1.884***												
					(0.222)												
Eastern Europe x Log WIID Gini index circa year 1980					1.205*												
					(0.712)												
Africa x Log Gini index 1700										0.000							
										(0.000)							
Asiax x Log Gini index 1700										0.000							
										(0.000)							
Western Europe x Log Gini index 1700										0.000							
										(0.000)							
Latin America x Log Gini index 1700										0.000							
										(0.000)							
North America x Log Gini index 1700										0.000							
										(0.000)							
Oceania x Log Gini index 1700										0.000							
										(0.000)							
Eastern Europe x Log Gini index 1700										0.000							
										(0.000)							
Africa x Log Gini index 1992															0.000		
															(0.000)		
Asiax x Log Gini index 1992															0.000		
															(0.000)		
Western Europe x Log Gini index 1992															0.000		
															(0.000)		
Latin America x Log Gini index 1992															0.000		
															(0.000)		
North America x Log Gini index 1992															0.000		
															(0.000)		
Oceania x Log Gini index 1992															0.000		
															(0.000)		
Eastern Europe x Log Gini index 1992															0.000		
															(0.000)		
Year fixed effects	No	No	Yes	No	No	No	No	Yes	No	No	No	No	Yes	No	No	No	No
Region fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	No	No	Yes	Yes	No	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No
Observations	660	311	660	311	311	311	224	660	311	311	311	224	660	311	311	311	224
Number of countries	44	39	44	39	39	39	28	44	39	39	39	28	44	39	39	39	28

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.13 Interaction between inequality and business environment, random effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Log Informal Economy (Hassan and Schneider)					Log Informal Economy (Medina and Schneider)				
LogGini index for year 1700	3.026*** (0.262)					6.011*** (0.225)				
LogGini index for year 1870		3.015*** (0.274)					5.985*** (0.225)			
LogGini index for year 1929			3.061*** (0.328)					4.564*** (0.283)		
LogGini index for year 1960				2.691*** (0.480)					1.496*** (0.261)	
LogGini index for year 1992					3.863*** (0.377)					1.418*** (0.279)
Getting credit score	-0.003 (0.005)	-0.003 (0.005)	-0.002 (0.005)	-0.002 (0.005)	-0.003 (0.005)	0.000 (0.002)	0.001 (0.002)	0.001 (0.003)	-0.000 (0.002)	-0.001 (0.003)
Cost of starting a business (% of income per capita)	-0.000 (0.002)	-0.000 (0.002)	-0.000 (0.002)	0.000 (0.002)	-0.000 (0.002)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)
Total tax rate (% of profit)	-0.005 (0.003)	-0.005 (0.003)	-0.006 (0.004)	-0.005 (0.004)	-0.005 (0.004)	-0.001 (0.004)	-0.001 (0.004)	-0.001 (0.005)	-0.001 (0.005)	-0.001 (0.005)
Getting credit scoret x LogGini index year 1700	-0.004 (0.005)					0.001 (0.003)				
Getting credit score x LogGini index year 1870		-0.004 (0.005)					0.002 (0.003)			
Getting credit scoret x LogGini index year 1929			-0.003 (0.006)					0.003 (0.003)		
Getting credit scoret x LogGini index year 1960				-0.003 (0.005)					0.001 (0.003)	
Getting credit scoret x LogGini index year 1992					-0.004 -0.006					-0.000 (0.003)
Cost of starting a business x LogGini index year 1700	-0.000 (0.002)					0.000 (0.001)				
Cost of starting a business x LogGini index year 1870		-0.000 (0.002)					0.000 (0.001)			
Cost of starting a business x LogGini index year 1929			-0.000 (0.002)					0.001 (0.001)		
Cost of starting a business x LogGini index year 1960				0.000 (0.002)					0.001 (0.001)	
Cost of starting a business x LogGini index year 1992					-0.000 (0.002)					0.001 (0.001)
Total tax rate (% of profit) x LogGini index year 1700	-0.005 (0.003)					-0.001 (0.003)				
Total tax rate (% of profit) x LogGini index year 1870		-0.005 (0.003)					-0.001 (0.004)			
Total tax rate (% of profit) x LogGini index year 1929			-0.006 (0.004)					-0.001 (0.005)		
Total tax rate (% of profit) x LogGini index year 1960				-0.006 (0.004)					-0.001 (0.005)	
Total tax rate (% of profit) x LogGini index year 1992					-0.006 (0.004)					-0.001 (0.005)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	970	970	970	970	970	1,117	1,117	1,117	1,117	1,117
Number of countrycode	123	123	123	123	123	137	137	137	137	137

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.14 Initial conditions and informal economy, random effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Log Informal Economy (Hassan and Schneider)					Log Informal Economy (Medina and Schneider)				
LogGini index for year 1700	1.048***					0.752***				
	(0.194)					(0.220)				
LogGini index for year 1870		1.048***					0.752***			
		(0.194)					(0.220)			
LogGini index for year 1929			1.428***					0.744***		
			(0.264)					(0.218)		
LogGini index for year 1960				1.030***					0.511***	
				(0.190)					(0.150)	
LogGini index for year 1992					0.983***					0.758***
					(0.182)					(0.222)
% Change in Gini indices for years 1700 and 1992	4.764***	4.764***	7.362***	7.139***	3.156***	-1.637***	-1.637***	-1.706***	-2.064***	-2.489***
	(0.979)	(0.979)	(0.990)	(0.980)	(1.084)	(0.199)	(0.199)	(0.196)	(0.211)	(0.285)
Getting credit score	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Cost of starting a business (% of income per capita)	0.001	0.001	0.001	0.001	0.001	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Total tax rate (% of profit)	-0.004	-0.004	-0.004	-0.004	-0.004	0.001	0.001	0.001	0.001	0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Log GDP per capita for year 1870	-1.657***	-1.657***	-1.657***	-1.657***	-1.657***	-1.564***	-1.564***	-1.564***	-1.564***	-1.564***
	(0.167)	(0.167)	(0.167)	(0.167)	(0.167)	(0.066)	(0.066)	(0.066)	(0.066)	(0.066)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	368	368	368	368	368	430	430	430	430	430
Number of countries	46	46	46	46	46	54	54	54	54	54

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.15 Club convergence classification using Hassan and Schneider's estimates

		Club 1 (18 countries)			
Peru	Thailand	The Gambia	Guinea	Benin	Liberia
Congo, Dem. R.	Tanzania	Burundi	Mozambique	Belize	Guatemala
Honduras	Bolivia	Uruguay	Cambodia	Lao PDR	Tajikistan

		Club 2 (30 countries)			
Russia	Greece	Spain	Argentina	Korea	Tunisia
Senegal	Mali	Burkina Faso	Niger	Sierra Leone	Ghana
Nigeria	Cameroon	Equatorial Guinea	Gabon	Madagascar	Comoros
Armenia	El Salvador	Nicaragua	Costa Rica	Haiti	Ecuador
Paraguay	Timor-Leste	Yemen, Rep.	Nepal	Azerbaijan	Dominican Republic

		Club 3 (28 countries)			
Egypt	South Africa	Italy	Mexico	Brazil	Venezuela
Malasya	Philippines	Vietnam	Turkey	Pakistan	Sri Lanka
Morocco	Algeria	Mauritania	Cote d'Ivoire	Togo	Congo, Rep. of
Rwanda	Uuganda	Malawi	Lesotho	Swaziland	Bulgaria
Guyana	Bangladesh	Bhutan	inidad and Tobago		

		Club 4 (31 countries)			
Netherlands	Beligum	Hungary	Romania	Denmark	Norway
Poland	Chile	Colombia	Japan	India	Mauritius
Central Afr. Rep.	Kenya	Zambia	Namibia	Botswana	Portugal
Ireland	Finland	Croatia	Slovenia	Macedonia	Slovak Republic
Suriname	Fiji	ong Kong SAR Chin	Lebanon	Unit. Arab Em.	Mongolia
The Bahamas					

		Club 5 (22 countries)			
France	United Kingdom	Sweden	Germany	Australia	Indonesia
New Zealand	Singapore	China	Iran	Canada	Angola
Guinea-Bissau	Luxemburg	Iceland	Malta	Czech Republic	Macao SAR China
Jordan	Saudi Arabia	Oman	Israel		

		Club 6 (7 countries)			
United States	Switzerland	Austria	Chad	Kuwait	Bahrain
Quatar					

A.1 Testing convergence

To analyse the transitional behaviour of the informal economy we use the convergence test proposed by Phillips and Sul (2007), also known as the log t regression. This test allow us to analyse the evolution of the size of the informal economy as a function of two components: one systematic permanent common component, g_{it}, and one transitory component, a_{it}.

$$Log(y_{it})= g_{it} +a_{it} \qquad (A.1)$$

where y_{it}, represents the size of the informal economy (as a percentage of GDP) for country i and time t. The systematic and transitory components can be re-expressed as shown in Eq. (A.2).

$$Log(y_{it})= \left(\frac{g_{it}+a_{it}}{u_t}\right) u_t = \delta_{it} u_t \qquad (A.2)$$

This equation expresses the informal economy, y_{it}, as a function of a common trend component, u_t, and an idiosyncratic component, δ_{it}. Phillips and Sul (2007) test for convergence by analysing this idiosyncratic component. To do so, they define a relative transition coefficient, h_{it}, of country i to the panel average at time t, as shown in Eq. (A.3).

$$h_{it} = \frac{log(y_{it})}{\frac{1}{N}\sum_{i=1}^{N} log(y_{it})} = \frac{\delta_{it}}{\frac{1}{N}\sum_{i=1}^{N}\delta_{it}} \qquad \text{(A.3)}$$

where h_{it} measures the idiosyncratic component δ_{it} relative to the panel average at time t. In other words, h_{it} measures the relative transition path of the size of the informal economy of country i relative to the panel average at time t, where $\delta_{it}=\delta_i + \sigma_i\xi_{it}L(t)^{-1} t^{-\alpha}$. That is, the time-varying systematic element δ_{it} is composed by a fixed parameter δ_i, ξ_{it} is iid(0,1) across i and weakly dependent over t, and varying function L(t) $\rightarrow\infty$ as t$\rightarrow\infty$. This formulation of the systematic element δ_{it} ensures that it converges over time to the fixed parameter δ_i for all $\alpha\geq0$. Moreover, this formulation allows for transitional heterogeneity and transitional divergence. Thus, by construction, the cross-sectional mean of h_{it} is unity, and the cross-sectional variance h_{it} converges to zero as shown in Eq. (A.4).

$$\sigma_t^2 = N^{-1}\sum_{i=1}^{N}(h_{it} - 1)^2 \rightarrow 0 \ as \ t \rightarrow \infty \qquad \text{(A.4)}$$

The property of Eq. (A.4) allows us to test the null hypothesis of convergence for all countries as Ho: $\delta_i=\delta$ and $\alpha\geq0$, against the alternative hypothesis of no convergence for some countries Ha: $\delta_i\neq\delta$ for some i and/or $\alpha<0$.

Rejecting convergence across the whole panel cannot rule out the existence of convergence across subgroups, or clubs, within the sample. Thus, Phillips and Sul (2007) develop a data-based algorithm which tests whether convergence is reached within subgroups of countries. We describe below the five steps involved in testing for club convergence, which we carry out using the logtreg stata package written by Du (2017).

Step 1. Cross-section ordering. The algorithm sorts countries in decreasing order according to the size of the informal economy in the last period or the average of the last half period. If there is substantial time-series volatility in the data, the sorting can be implemented based on the time series of the last fraction, such as half or a third of the sample.

Step 2. Identify the primary core group of *k** countries. The algorithm then selects the first *k* countries with the largest size of the informal economy in the panel to form the subgroup G_k for *N> k> 2* countries and estimate the convergence t-test for the subgroup. If there are no *k* countries satisfying such restriction, the algorithm continues adding countries of core group size *k** until the t-test> -1.65 concluding there is convergence. If there is a single convergence group including all countries, then the size of the club is *N*.

Step 3. Sieve the data for new club members. The algorithm continues by adding another country at a time to the core primary group with *k** members and runs the log t test again. This new country is included in the subgroup convergence club if the corresponding convergence t-test>-1.65.

Step 4. Recursion and stopping rule. Form a subgroup of countries among those not sieved in step 3. If the corresponding convergence t-test>-1.65, the subgroup forms another convergence club. Otherwise, the algorithm repeats steps 1-3 on this subgroup.

Step 5. Club merging. New club classifications are obtained by merging the countries composing the initial clubs. Clubs 1 and 2 are merged and then run the log t test to assess whether they fulfil the convergence hypothesis jointly. If so, these two clubs form a new Club 1. Then this new Club 1 is merged with the initial Club 3 if they do converge, the process carries adding another club. If new Club 1 and initial Club 3 do not converge, then the initial Club 2 is merged with Club 3, so on and so forth until no new clubs can be merged any more.