Inequality and informality revisited

Verónica Amarante* and Rodrigo Arim**

Abstract

Many developing economies share two distinctive features: high levels of inequality and sizeable informal employment. Despite the importance of these facts, the literature addressing the relationship between both variables is scarce. In this paper, we propose a simple model which illustrates their ambiguous relationship: an increase in the level of formalization in the labor market could have different effects on income inequality. According to this model, different factors, such as the wage differential between formal and informal sectors, the level of inequality within each sector and the size of the informal sector may affect the sign of the relationship between inequality and informality. We empirically test this hypothesis using panel data for Latin America countries, covering the period 1990-2014. Our results indicate that, for the specific conditions that held on that region and period, increases in formalization in the labor market have been related to decreases in wage inequality.

Keywords: informality, inequality, Latin America
JEL Code: D31, J31

Introduction

Most developing economies are characterized by sizeable informal employment (see for example Schneider, 2012; Tornarolli et al, 2014; La Porta and Shleifer, 2014). In many of these economies, the levels of income inequality are also substantial, and mainly determined by high labor income inequality. Despite these two distinct facts, the literature addressing the relationship between informality and inequality is relatively scarce. This is probably related to the complexity of this relationship. In effect, one may address the issue considering that income inequality is a determinant of informality, or argue in favour of the opposite causation, claiming that informality is the origin of inequality. An example of the first perspective is the work by Chong & Gradstein (2007), who present a model where increases in income inequality cause a growth of the informal sector by lowering the relative benefits of formality for the poor. This effect becomes larger when institutions are weak. On the same line, Mishra and Ray (2014) have argued that higher income inequality determines higher demand for informal sector goods, providing evidence of this positive relationship for developing countries. Related

---

* The authors are thankful to Martín Brun for excellent research assistance.
* Economic Commission for Latin America and the Caribbean, Chief of the Office in Montevideo.
** Universidad de la República, Facultad de Ciencias Económicas y de Administración, Instituto de Economía.
literature emphasizes that high informality implies lower tax revenue (Loayza, 1996; Johnson, Kaufmann, and Zoido-Lobatón, 1998; Schneider and Enste, 1998; among others). This in turn can affect State capacity to redistribute, both directly and indirectly through education, and thus has a direct impact on inequality. This type of argument has been incorporated by Rosser et al. (2000, 2003), who also find a positive correlation between informality and the Gini coefficient in transition economies, and explicitly recognize that different causal mechanisms may operate in both directions. Attanasio and Binelli (2010), using microdata for Mexico, show that wage inequality is much higher among informal workers and that the changes in the size of the informal sector closely follow changes in wage inequality. In a later paper, Binelli (2016), also based on Mexican evidence, argues that higher wage dispersion is one of the channels through which informality negatively affects development. More recently, Amarante et al (2016) analyze wage inequality dynamics in Uruguay using micro econometric decompositions, and their results highlight the importance of the formalization process in the labor market as a key factor in understanding the decrease in earnings inequality. While the authors acknowledge that informality cannot be claimed to be exogenous, they argue that the effect of this variable can be interpreted as the effect of labor market institutions. The regulatory framework within which labor relations develop in the formal economy – e.g. mechanisms to determine wages, regulations of minimum wages, social security benefits, etc. – highly determines the observed decline in wage inequality during the period.

The link between inequality and informality, which has been scarcely addressed, is not an easy one given the unclear direction of causality and feedback between both phenomena. There seems to be, however, some evidence of a positive correlation between inequality and informality in some specific country studies in the region. What is clear is that methodological problems when addressing this relationship are relevant. Specifically, problems related to omitted variables, reflecting diverse social and economic interventions which can influence both inequality and informality, cannot be ignored. Keeping these limitations in mind, this article attempts to contribute to the discussion about this relationship, by first providing a simple analytical model that illustrates the nature of the relationship between inequality and informality (section 1). Based on panel data for Latin American countries for the period 1990 to 2014 (see section 2), we provide new empirical evidence on the relationship between informality and inequality (section 3). We finally present our concluding remarks (section 4).

1. Informality and inequality: a basic model

A basic analytic framework may be useful in order to shed some light on the difficulties associated with identifying the impact that labor formalization processes - such as the ones experienced in the last decade in Latin America - have over inequality. Up to now, we have referred to inequality in a broad sense, but it is necessary to make some precisions. The model we are presenting refers to wage inequality, as it’s the main driver of total household inequality in developing economies. Our model is an adaptation of the one proposed by Anand & Kanbur (1993) to explain the relationship between development and inequality. We consider an economy in which each individual works either on the formal sector (sector a) or the informal sector (sector b). In a given moment, $x(t)$ is the share of workers in the formal sector and $[(1 - x(t))]$ is the share occupied by the informal sector\(^1\). The informal sector is characterized by lower productivity and, hence, lower wages, as well as a non-wage benefits scheme – defined work schedule, labor rights, social security coverage, etc. - less desirable for

\(^1\)This assumption excludes the possibility of mixed labor participation, with people occupied partially in the formal sector and partially by the informal sector.
the employee. This last assumption is not unequivocally clear. It has been pointed out that in some cases informality could reflect a rational ‘escape’ strategy of the employees and employers, who could value the benefits of informal work (present consumption, flexibility) more than the benefits resulting from formalization. This could be reinforced by a State which is highly inefficient in its role as a provider of benefits and as regulator. The second perspective highlights the «exclusion» process and considers informal employees as victims of a dual and segmented system that fails to offer access to social protection. In the analysis of informality in Latin America presented in Perry et al (2007), which discusses both perspectives, workers are characterized according to their motivations and preferences for non-monetary social benefits (flexibility, stability, etc.). According to the authors, motivations differ significantly amid groups. Most of the informal independent workers (self-employed and employers) would have entered voluntarily in this category in order to escape from the formal social protection systems. Their informal status obeys to a rational election, associated to some characteristics of the informal job positions that make them more desirable in certain contexts and conditions. On the other hand, there would be employees remaining outside formal contracts and social protection institutions who wouldn’t have chosen to be there; their situation would be the result of exclusion, not of a rational decision.

The formalization process implies a trend increase of \( x(t) \). Let \( F_a(y, x(t), t) \) and \( F_b(y, x(t), t) \) be the distribution functions of wages \( y \) in the formal and informal sectors of the economy respectively. Since \( a \) and \( b \) represent an exhaustive share of the workers, the distribution and density functions of wage are a weighted average of the sector’s distributions:

\[
F(y, x(t), t) = x(t)F_a(y, x(t), t) + \left[ (1 - x(t)) \right] F_b(y, x(t), t) \quad (1.1)
\]

\[
f(y, x(t), t) = x(t)f_a(y, x(t), t) + \left[ (1 - x(t)) \right] f_b(y, x(t), t) \quad (1.2)
\]

Expressions (1.1) and (1.2) are just identities, but they capture the complexity of the link between inequality and the transition dynamic between the formal and informal sectors of the economy. Let \( I_t[F(y, x(t), t)] \) be a synthetic index that characterizes inequality in moment \( t \). See that \( I_t \) varies as a function of the changes in the relative share of the formal sector \( x(t) \) and the changes in the distribution functions of each sector \( F_a \) and \( F_b \). As economic development unravels, it is expectable that \( x(t) \) rises, but the distributions \( F_a \) and \( F_b \) also change prompted by diverse factors that usually arise along economic growth: changes in labor participation rates, in demographic dynamics and in schooling returns, among others. Moreover, it is expectable that the impact of these factors in both distributions turn out to be deeply asymmetric. For instance, if there is a persistent gap favoring the formal sector (even after controlling by a set of relevant personal characteristics) and if the transition from the informal sector towards the formal sector implies a selective process in which the most productive workers from the informal sector are the ones able to transfer towards the formal sector first (which seems reasonable), inequality within the informal sector would decrease as the group with higher income disappears while inequality within the formal sector rises due to the increase of workers in the lower part of the distribution coming from the informal sector. The former indicates that inequality within the formal sector and the informal sector \( (I_a \neq I_b) \) is not independent from \( x(t) \).

In turn, \( F_a \) and \( F_b \) can evolve differently due to reasons unrelated to \( x(t) \). For example, if the distribution of workers regarding their educational achievements systematically differs between the formal sector and the informal sector, changes in schooling returns would have different distributional impacts amid groups. Therefore, from a theoretical point of view,
isolating and identifying the specific impact of formalization –understood as an increase in $x(t)$ - over wage distribution is complicated. A first approximation could be obtained if we suppose that the distribution functions of both sectors remain unchanged and are not dependent of their relative share on total employment. In that case, $F_a$ and $F_b$ are independent from $x(t)$ and $t$. Following Anand & Kanbur (1993), the expression (1.1) could be expressed as follows:

$$F(y, x(t)) = x(t)F_a(y) + [(1 - x(t))]F_b(y)$$

$$f(y, x(t)) = x(t)f_a(y) + [(1 - x(t))]f_b(y)$$

Wage distributions in the formal and informal sector are assumed constant throughout time, with both the average income and inequality level remaining unchanged. In addition, the fact that $F_a$ and $F_b$ are independent from $x(t)$ implies that those who transfer from the informal sector to the formal sector are randomly selected, without changing the distribution of the original sector and the sector of destination.

The presence of higher levels of productivity in the formal sector implies higher average wages. The expected relationship regarding the inequality between both sectors is a priori less clear. However, the usual assumption is that inequality is greater in the formal sector, being understood that there is a greater dispersion of productivity levels and job characteristics. In contrast, the informal sector is characterized by lower productivity dispersion, strongly concentrated on reduced levels. These assumptions can be formalized as follows:

$$\mu_{a,t} > \mu_{b,t}$$

$$I_{a,t} > I_{b,t}$$

Being $\mu_{i,t}$ and $I_{i,t}$, respectively, the mean wage and a generic inequality index. In this case, $p$ stands for:

$$p = F(z, x) = xF_a(z) + (1 - x)F_b(z)$$

The Lorenz curve, which captures the share of income perceived by the population below the $p^{th}$ percentile, can be define as a function of the relative share of the formal sector and the informal sector in total employment:

$$L(p, x) = \frac{1}{\mu} \int_{0}^{x(z, p, x)} yf(y)dy =$$

$$= \frac{1}{\mu} \left\{ x \int_{0}^{x(z, p, x)} yf_a(y)dy + (1 - x) \int_{0}^{x(z, p, x)} yf_b(y)dy \right\} =$$

$$\frac{x}{\mu}L_a(F_a(z)) + \frac{(1 - x)\mu_b}{\mu}L_b(F_b(z))$$

The Lorenz curve for the whole economy is the average of the Lorenz curve for each sector ($L_a(F_a(z))$ and $L_b(F_b(z))$), weighted by the relative share of the formal sector and the informal sector in the employment.

The expression (1.6) shows that the Lorenz curve valued in the $p^{th}$ share of the occupied depends on three factors:

1. **Sector’s distribution of the employment.** The expressions $x$ and $(1-x)$ measure the relative share of the formal sector and the informal sector in total employment.
2. **Intra-sector inequality.** This factor is captured by \( z (L_a(F_a(z))) \) and \( z(L_b(F_b(z))) \), which measure the share of total wages from each sector that is earned by the occupied who earn less than \( z \).

3. **Average wage differential.** Relationship between the average wages of the formal and the informal sector regarding total average wage (\( \frac{\mu_a}{\mu} \) and \( \frac{\mu_b}{\mu} \)).

Changes in the level of labor formalization won’t have an impact exclusively on a sector’s employment distribution, but also, through multiple channels, on its inequality. In fact, unless the transition from the informal sector towards the formal sector is completely random, the transfer would change both the distribution of the formal sector and the informal sector, therefore impacting intra-sector inequality. Likewise, average wage differential would vary prompted by the changes in the sectors’ integrations.

In order to analyze the impact of an increase in the formalization level over inequality, it is useful to consider the interaction between these three factors, as seen in expression (1.6). According to the Lorenz dominance criteria\(^2\), in order for the distribution in moment \( t \) to be considered more equitative than the one in moment \( s \) it must be met that \( L(p, x_t) \geq L(p, x_s) \) for all \( p \). In other terms, the Lorenz curve for moment \( t \) must be systematically above the Lorenz curve for moment \( s \\(^3\).

Let \( x^1 \) and \( x^2 \) be the share of formal workers in the years 1 and 2 respectively, being \( x^2 > x^1 \). The difference between the Lorenz curves for years 2 and 1 is:

\[
L(p, x^2) - L(p, x^1) = \left( \frac{x^2 \mu_a^2}{\mu^2} L_a^2(F_a(z_2)) + \frac{(1-x^2) \mu_b^2}{\mu^2} L_b^2(F_b(z_2)) \right) - \left( \frac{x^1 \mu_a^1}{\mu^1} L_a^1(F_a(z_1)) + \frac{(1-x^1) \mu_b^1}{\mu^1} L_b^1(F_b(z_1)) \right) \tag{1.7}
\]

It is not possible a priori to identify a direct relationship between the formalization increase and inequality. Thus, it appears necessary to incorporate some additional structure.

Using the usual wage distribution decomposition approach, it is possible to decompose the total change in the distribution in three factors, based on the identification of two sets of counterfactual statistics. First, we define the counterfactual that characterizes the situation in which the changes in the formalization level will operate without variations in the mean wages and the inequality within sectors (counterfactual distribution 1). Secondly, we calculate the counterfactual distribution in which mean wages of each sector change as observed between year 1 and 2, but inequality within formal sector and informal sector remain unchanged (counterfactual distribution 2).

\(^2\)The main inequality indexes used in the related literature meet Lorenz dominance criteria.

\(^3\)Lorenz dominance is a demanding criterion, as it fails to allow the classification of situations in which curves corresponding to different moments intersect. However, for that former reason it constitutes a solid starting point in order to classify the distributive impact of changes in formality, as the analysis is independent from the particularities of the different families of inequality synthetic indexes.
Let \( L^{1,x^2}(x_2,p) \), the Lorenz curve in moment 1 if the relative share of the formal sector matches the share observed in year 2, be the curve representing the counterfactual distribution 1. This transformation simulates the curve with the assumption that the distribution of remunerations within the sector is the one observed in moment 1, with their respective mean and dispersion, but the distribution of employment amid sector is the one corresponding to moment 2.

The second counterfactual distribution is defined by \( L^{1,x^2,\mu_1^2}(x_2,p) \), which represents the value of the Lorenz curve in the \( p^{th} \) percentile if the distribution of employment amid sectors and the mean wages of each sector in moment 2 is imposed over the distribution of year 1.

Based on these distributions, total change of Lorenz curves for each percentile \( p \) can be defined as following:

\[
L(p, x^2) - L(p, x^1) = \frac{L^{1,x^2}(x_2,p) - L(p, x^1)}{\text{composition effect}} + \frac{L^{1,x^2,\mu_1^2}(x^2,p) - L^{1,x_2}(x^2,p)}{\text{wage structure effect}} + \frac{L(p, x^2) - L^{1,x^2,\mu_1^2}(x_2,p)}{\text{intra-sector inequality effect}}
\] (1.8)

Take into consideration that the decomposition acts by the successive addition and subtraction of the counterfactual distributions. The final result is the difference between the distribution of years 1 and 2, which represents the change effectively observed. Each component of the decomposition corresponds to one of the three factors on which the Lorenz curve depends, as commented based on the expression (1.6). Following, we analyze the derivation of the counterfactual distributions and the identification of the three effects.

**Counterfactual distribution 1. Increase in formality without changes in the sector’s distribution (\( \mu_i^2 = \mu_i^1 ; F^1_i(z) = F^1_i(z) \) for all \( z, i=\text{a,b} \)).** Under this configuration, changes in formalization impact the distribution exclusively through the variations in the composition of employment, without changing the prevailing distribution in the formal and informal sector. It is assumed that mean wages and dispersion within both sectors are independent from changes in \( x \). In this case, based on the expression (1.6), the first counterfactual distribution is defined as:

\[
L^{1,x^2}(x^2,p) = \frac{x^2 \mu_a^1}{\mu_1 x^2} L^1_a \left( F^1_a \left( z^{1,x^2} \right) \right) + \frac{(1-x^2) \mu_b^1}{\mu_1 x^2} L^1_b \left( F^1_b \left( z^{1,x^2} \right) \right)
\] (1.9)

Factor \( z^{1,x^2} \) represents the wage’s threshold that defines percentile \( p \) in moment 1 if the share of workers in the formal sector changes from \( x^1 \) to \( x^2 \). This counterfactual distribution has no impact on individual sector’s distributive statistics, as by hypothesis the distributions of the formal and informal sector remain unchanged and in consequence neither the mean wages nor the Lorenz curves of each sector show variations. Notwithstanding the latter, the overall distribution is changed, prompted by the variation in the relative share of the formal sector (\( x^2 \)).

Particularly, the mean wage and the threshold defining each percentile \( p \) are modified:
Given that the mean remuneration in the formal sector is greater than the one in the informal sector, the impact of formalization—without variations in the sector’s distribution over mean wages is clearly positive. On the other hand, the threshold $z$ is higher under the new configuration, as a larger share of the wage earner population is occupied by the higher-wage sector ($z^{1,x^2} \geq z^1$).

Despite the latter, it is not possible to reach similar conclusions regarding other statistics, particularly in those capturing the dispersion. The composition effect, defined as in (1.8), highlights this ambiguity. In order to shed some light over the situation we incorporate further elements. The change in the mean wage fostered by the variation in the participation in the formal sector is denominated $\beta$, such that $\mu^{1,x^2} = (1 + \beta)\mu^1$. Based on this notation, we derived the composition effect in the Annex, which is defined as follows:\(^5\)

$$\text{Composition effect} = L_i^{1,x^2}(x_2, p) - L(p, x_1) = \frac{L_b(z^1) + x^2 L_a(z^{1,x^2}) - x^1 (1 + \beta) L_a(z^1)}{L G L_a(z^1) + (1 - x^2) L_b(z^{1,x^2}) - (1 - x^1)(1 + \beta) L_b(z^1)}$$

The first factor is always positive and captures the relative distance between the distributions of the formal and informal sector based on the relationship between the Generalized Lorenz curves of each sector\(^6\). Notice that the greater the gap between both distributions, the lesser the redistributive effect of an increase in formality. The idea is that the remaining workers in the informal sector grow apart from the expanding formal sector at increasing rates. In consequence, the positive effect of higher formality is buffered by the greater relative inequality within the workers that remain in the informal sector.

The second term, which basically captures the intra-sector differences in inequality, is ambiguous both in sign and absolute value. The ratio measures the relationship between the changes in the Lorenz curves of the formal sector and the informal sector\(^7\), weighted by the variation in the relative participation of each sector and the increase in the overall mean wage. The denominator and numerator are approximations to the degree of internal inequality in the formal and informal sector respectively, the difference between \(L_i(z^{1,x^2})\) and \(L_i(z^1)\) representing the bulk of total remunerations generated by the \(i^{th}\) sector which is appropriated by the population belonging between the percentiles constrained by and \(z^{1,x^2}\) and \(z^1\). Notice that this term will tend to be positive and show a higher absolute value the greater the change

\(^4\) We assume that the distribution function of the formal sector stochastically dominates the distribution function of the informal sector, that is to say that $F_a(z) \leq F_b(z)$ for all $z$. Therefore, it stands that $F_a(z) \leq p \leq F_b(z)$. If the relative share of the formal sector increases, $z$ must rise in order that $p$ remains unchanged.

\(^5\) For simplicity, $L_i(F_i(z))$, is written as $L_i(z)$, as there is no ambiguity in the interpretation. $L(z)$ is the share of wages in sector $i$ that are appropriated by the $p^{th}$ percentile as defined by threshold $z$, $p_i = F_i(z)$.

\(^6\) The Generalized Lorenz curve is defined as the product of the Lorenz curve and the mean wage.

\(^7\) Lorenz curves corresponding to each sector remain unchanged. In turn, $z$, the thresholds used to determine percentiles $p$ in the distribution, change.
in formality \((x^1 \rightarrow x^2)\) is and the greater the change in the formal sector Lorenz curve is in comparison to the one registered in the informal sector. However, the final impact will depend on the relative position of the Lorenz curves. This result reveals the importance of the difference between inequalities in both sectors in order to determine the final impact in overall inequality.

In other words, if the differential favoring the distribution of the formal sector \(\rightarrow\) towards where workers transfer is not compensated by changes in the differences in the inequality within the sectors, then an increase in the participation in the formal sector leads to a fall in the dispersion of wages.

**Counterfactual distribution 2. Increase in formality without changes in the wage dispersion within sectors**, but with increases \(\alpha_a\) and \(\alpha_b\) in the mean wages of the formal and informal sector respectively \((\mu_a^2 = \varphi_1 \mu_1^2 ; L_1^i(p)=L_1^i(p)\) for all \(i=a,b\), being \(p\) the percentiles in the distribution). Deriving the second counterfactual distribution implies the imposition of the formal sector share and the mean wages per sector corresponding to year 2 onto the distribution of year 1. The specific transformation of the distribution function, the Lorenz curve, is invariant.

In the Annex we derivate the wage structure effect, as defined in the equation (1.8). Its expression is as follows:

\[
Wage\ structure\ effect = L^{1,x_2,\mu_1^2}(x_2,p) - L^{1,x_2}(x_2,p) \\
= x^2 \left[ \frac{(1 + \varphi_a)L_a^1(z^2)}{(1 + \alpha)} - L_a^1(z^1) \right] \\
+ (1 - x^2) \left[ \frac{(1 + \varphi_b)L_b^1(z^2)}{(1 + \alpha)} - L_b^1(z^1) \right] 
\]

Wage structure effect is the average \(\rightarrow\)-weighted by the share of each sector in year 2 of the changes in the Lorenz curves of the formal and informal sector. The curves corresponding to year 2 are adjusted by the change in the wage structure, measured through the growth rate of the mean wage of each sector.

If the gap between the formal sector and the informal sector narrows \(\varphi_b > \alpha > \varphi_a\), then the second term is unequivocally positive, although the first term could flip this deconcentrating impact if the change in the formal sector Lorenz curve is not high enough to overshadow the drop in the formality premium. The opposite would occur if the gap widens \(\varphi_b < \alpha < \varphi_a\), as the positive effect over inequality of the first term could be overtaken if the change in the formal sector Lorenz curve is lower than the fall in the wage gap. Once again, the average wage premium dynamic has an ambiguous impact on inequality, as it is boosted or constrained depending on the changes on inequality within the sectors, measured through its respective Lorenz curves.

Finally, the effect of the changes on inequality within sectors is resumed by the following equation:

\[
L^2(x^2,p) - L^{1,x_2,\mu_1^2}(x^2,p) \\
= x^2 \mu_a^2 \left[ L_a^2(z^2) - L_a^1(z^1) \right] \left[ 1 - x^2 \right] \mu_i^2 \left[ L_b^2(z^2) - L_b^1(z^1) \right] 
\]

(1.14)
In this case, the impact would be deconcentrating if the following condition is met:

\[
\frac{x^2 \mu_a^2}{(1 - x^2) \mu_b^2} \geq \frac{[L_b^b(z^2) - L_b^b(z^1)]}{[L_a^a(z^2) - L_a^a(z^1)]}
\] (1.15)

The relationship between the changes of the Lorenz curves of the formal and informal sector must be lower than the wage gap weighted by the relative share of each sector in total employment. The probability of this happening is increasing with \(x^2\) and with the size of the wage gap.

The three terms of the decomposition show ambiguous signs. The straightforward conclusion of the former exercises is that the direction and magnitude of the changes are not easily predictable. The potential effects on wage inequality of an increase in the size of the formal sector depend on the wage differential between the formal and informal sector and also on the within levels of inequality in both sectors.

2. Specification and data

To explore the relationship between informality and inequality taking into account the implications from the model developed in the previous section, we estimate a basic reduced form equation - standard in the literature since the seminal papers of Deininger and Squire (1998) and Barro (2000, updated 2008):

\[
G_{it} = \alpha + \beta_1 \ln f_{it} + \beta_2 \frac{MW_{f_{it}}}{MW_{it}} + \beta_3 \frac{GW_{f_{it}}}{GW_{it}} + \beta_4 GDP_{pc_{it}} + \beta_j X_{it} + \epsilon_{it}
\] (1)

where inequality in labor income \((G_a)\) is related to the size of the informal sector (informality rate, \(lnf_a\)), the wage differential between the formal and the informal sector (ratio between mean wage in the formal and mean wage in the informal sector, \(\frac{MW_{f_{it}}}{MW_{it}}\)) and the internal levels of inequality in both the formal and informal sectors (ratio between the Gini of wages in formal and informal sectors, \(\frac{GW_{f_{it}}}{GW_{it}}\)). We include the level of GDP per capita \((GDP_{pc_{it}})\) and a set of control variables (reflected by \(X_{it}\)).

An estimation of the causal effect of a decrease in informality on inequality (given by the coefficient \(\beta_1\)) would require an exogenous source of variation for each of the variables that are included in the regression, since all these factors are arguably endogenous to income inequality, as discussed in the previous section. This is a relevant shortcoming from our approach and should be considered when interpreting the estimated coefficients.

Our estimations are based on an unbalanced dataset of 18 Latin American countries over the years 1990-2015. We combine variables from different datasets to construct this panel data set. Our main dependent variable reflects labor income inequality, and comes from the Socio-Economic Database for Latin America and the Caribbean (SEDLAS), compiled by CEDLAS and
The World Bank.\textsuperscript{8} It is measured as net labor income (after taxes). The size of the informal sector is measured as the share of salaried workers with no right to pensions when retired, and also comes from SEDLAS database.

The other two relevant dependent variables, mean formal and informal wages and inequality for formal and informal wages, were calculated from annual data coming from national household surveys. In both cases, formal workers are those with right to pension when retired in their main job.\textsuperscript{9}

Finally, we also control for other variables traditionally used to explain inequality. We consider GDP per capita in constant USD dollars from the World Economic Outlook of the International Monetary Fund, Schooling (average completed years for people aged 25-65, taken from the Socio-Economic Database for Latin America and the Caribbean (CEDLAS and The World Bank). Table 2 provides summary statistics for the variables used in this article.\textsuperscript{10}

\begin{table}[h]
\centering
\begin{tabular}{lcccc}
\hline
Variable & Obs & Mean & Std. Dev. & Min & Max \\
\hline
\textbf{Main Independent Variables} & & & & & \\
Gini (wage) & 292 & 0.495 & 0.048 & 0.382 & 0.607 \\
\hline
\textbf{Main Dependent Variables} & & & & & \\
Informality rate & 220 & 0.455 & 0.172 & 0.125 & 0.773 \\
Mean Formal Wage & 147 & 139.0 & 19.9 & 106.4 & 217.2 \\
Mean Informal Wage & 147 & 63.4 & 12.5 & 42.7 & 85.8 \\
Gini Formal Wage & 147 & 0.407 & 0.060 & 0.268 & 0.529 \\
Gini Informal Wage & 147 & 0.475 & 0.083 & 0.338 & 0.692 \\
\hline
\textbf{Control Variables} & & & & & \\
GDP per capita (log) & 464 & 8.13 & 0.757 & 6.58 & 9.74 \\
Unemployment rate & 445 & 7.80 & 3.962 & 1.30 & 20.70 \\
Schooling (years) & 296 & 7.6 & 2.111 & 0.0 & 11.1 \\
Fertility & 450 & 2.94 & 0.797 & 1.76 & 5.58 \\
Gov. expenditure & 449 & 12.0 & 3.905 & 3.0 & 43.5 \\
\hline
\end{tabular}
\caption{Summary Statistics (variables in levels)}
\end{table}

Source: own estimations

The positive correlation between inequality and informality found in previous research seems also to be present in Latin American countries. As Figure 1 illustrates, both pooled data and—in most cases—countries’ experiences, indicate that the level of wage inequality is positively associated with informality, measured as the percentage of workers who do not contribute to

\textsuperscript{8} Details can be found at http://sedlac.econo.unlp.edu.ar/
\textsuperscript{9} This information is not always available for all workers. A dummy variable is included in the database to signal those cases where the variable only refers to salaried workers.
\textsuperscript{10} An overview of the definition and sources of variables used in this article can be found in table A.1.
the social security. A more detailed analysis of this relationship is presented in the following section.

Figure 1. Wage Inequality and informality in Latin America Latina. 1990-2014

a. All countries  

b. By country

Source: based on SEDLAC

3. Results

The high levels of inequality in labor income and the importance of informality, understood as non-registered work, are two distinctive characteristics of labor markets in the region. In consequence, there is a vast body of literature analyzing the determinants of labor income inequality and its evolution, which mainly focuses in the relative demand for qualification in order to explain recent changes (see Lustig & López Calva, 2010; Gasparini et al, 2012; Azevedo et al, 2013; among others), although institutional framework has also been considered (see ECLAC, 2014). On the other hand, the analysis of informality determinants in the region has also been addressed (see Perry et al, 2007; ECLAC, 2013; among others), as the lack of coverage of social security affects around 55% of Latin American employees (ECLAC, 2013). Surprisingly, informality has not been considered a relevant factor in econometric estimations trying to explain variations in inequality between countries or in time, as discussed in section 1. Of course, this may be in part explained by the difficulties related to the availability of data. We try to fill this gap by providing new evidence on the relationship between inequality and informality, using a panel dataset of 18 Latin American countries over the years 1990-2015.

As a first step we analyze the correlations between the main variables in our model. Wage inequality as measure by the Gini index, the informality rate and the wage gap between the formal and the informal sector show a positive but moderate correlation in our panel data set of Latin American countries (see Table 2). The absolute values of the correlation coefficients are relatively low, suggesting that the links between these dimensions of the regional labor markets are weak. The exception is intra sector inequality, measured by the ratio of the Gini index of wages in the formal sector and the Gini index of wages in the informal sector, which is
negatively correlated with the informality rate, although the magnitude of the coefficient is small.

Table 2. Simple correlation matrix

<table>
<thead>
<tr>
<th></th>
<th>Wage Inequality (Gini)</th>
<th>Informality rate</th>
<th>Wage gap : formal vs informal</th>
<th>Intra sector inequality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wage Inequality (Gini)</td>
<td>1</td>
<td>0.3483</td>
<td>0.1981</td>
<td>0.3655</td>
</tr>
<tr>
<td>Informality rate</td>
<td>0.3483</td>
<td>1</td>
<td>0.0998</td>
<td>0.0431</td>
</tr>
<tr>
<td>Wage gap : formal vs inf</td>
<td>0.1981</td>
<td>0.0998</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Intra sector inequality</td>
<td>0.3655</td>
<td>-0.0431</td>
<td>0.036</td>
<td>1</td>
</tr>
</tbody>
</table>

Source: own estimations

We regress wage inequality, as measured by the Gini index, against the wage gap between the formal and the informal sector, the informality rate, the intrasector wage inequality and an interaction between the informality rate and the wage gap, controlling for GDP per capita (in logs, PPP). The first three variables are derived directly from the model presented in section 1. We incorporated an interaction term between informality and the average wage gap because the theoretical discussion of previous section suggests that the effect of both dimensions depends on the level of the other. Our data set comprises 16 Latin American countries for the period 1990-2013.

Pooled OLS estimations show no significant coefficients for most of the interest variables. Nevertheless, when the estimations consider the panel structure of data, both the informality rate and the average wage gap emerge as strongly significant (table 3). Also, the interaction term becomes relevant. These results would imply that both an increase in incidence informality or in the wage gap between sectors lead to variations in the same direction of wage inequality, but the magnitude of the effect varies according the level of both variables. These results emerge with fixed effects estimations and with random effects estimations. Under GMM estimations, the signs are similar but the significance is weaker. The levels of wage inequality within each sector (as measured by the ratios of the Gini coefficients in our specifications) are not significant in our estimations.


<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1) pooled OLS</th>
<th>(2) Fixed Effects</th>
<th>(3) Random Effects</th>
<th>(4) Arellano-Bond</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gini = Lag</td>
<td></td>
<td></td>
<td>0.349 [0.0942]***</td>
<td></td>
</tr>
<tr>
<td>Informality rate</td>
<td>0.0332</td>
<td>0.315</td>
<td>0.305</td>
<td>0.209</td>
</tr>
</tbody>
</table>
The results are robust to the inclusion of a set of control variables potentially relevant to explain the difference in wage inequality among Latin American and Caribbean countries (table 4). These include unemployment rate, education, measured by average years of schooling, public expenditure, proxied by general government final consumption expenditure, and fertility rate. As is usual in the literature with panel of countries, we opted for the fixed effects estimator. Again, wage inequality is positively related to the level of informality in the labor market, as well as the average wage gap between the formal and the informal sector. The interaction term keeps being negative and weakly significant, indicating that the positive association between inequality and both the wage gap and the informality rate depends on the levels of each.

The results are robust to the inclusion of a set of control variables potentially relevant to explain the difference in wage inequality among Latin American and Caribbean countries (table 4). These include unemployment rate, education, measured by average years of schooling, public expenditure, proxied by general government final consumption expenditure, and fertility rate. As is usual in the literature with panel of countries, we opted for the fixed effects estimator. Again, wage inequality is positively related to the level of informality in the labor market, as well as the average wage gap between the formal and the informal sector. The interaction term keeps being negative and weakly significant, indicating that the positive association between inequality and both the wage gap and the informality rate depends on the levels of each.

The results are robust to the inclusion of a set of control variables potentially relevant to explain the difference in wage inequality among Latin American and Caribbean countries (table 4). These include unemployment rate, education, measured by average years of schooling, public expenditure, proxied by general government final consumption expenditure, and fertility rate. As is usual in the literature with panel of countries, we opted for the fixed effects estimator. Again, wage inequality is positively related to the level of informality in the labor market, as well as the average wage gap between the formal and the informal sector. The interaction term keeps being negative and weakly significant, indicating that the positive association between inequality and both the wage gap and the informality rate depends on the levels of each.

The results are robust to the inclusion of a set of control variables potentially relevant to explain the difference in wage inequality among Latin American and Caribbean countries (table 4). These include unemployment rate, education, measured by average years of schooling, public expenditure, proxied by general government final consumption expenditure, and fertility rate. As is usual in the literature with panel of countries, we opted for the fixed effects estimator. Again, wage inequality is positively related to the level of informality in the labor market, as well as the average wage gap between the formal and the informal sector. The interaction term keeps being negative and weakly significant, indicating that the positive association between inequality and both the wage gap and the informality rate depends on the levels of each.
Interestingly, aggregated control variables reinforce the results. The three dimensions of informality that we identified as possible mechanisms that could incise on wage inequality – prevalence of informality, sectorial wage gap, and difference in the inequality between sectors – appear to have significant effects. The implicit magnitudes of the effects are relatively important. To indicate an order of magnitude, for example, our estimations point out that a variation of one percentage point in the prevalence of informality is associated with a change of practically 0.3 points of the Gini index in the same direction. The interaction term shows that this impact is enhanced the greater the wage differential. In turn, given the same informality rate and the level of the control variables, an average salary differential ten percent greater causes an increase of approximately 0.8 points of the Gini. At the same time, the differences in the extent of internal inequality in the sectors are relevant to explain the general wage inequality.

Therefore, our main contribution to the literature about wage inequality in Latin American and Caribbean lies in underlying the importance of institutional norms that characterize the labor markets in the region. Probable the most important of them is the persistent level of informality and the difference in the wage setting mechanisms in both sectors. We introduced this feature in the regression analysis that tries to identify the determinants of inequality, and found significant effects of informality and other related variables. Further research is needed to understand how norms and institutions that delineate the functioning of formal and informal labor market affect wage inequality.

### 4. Final comments

Although it isn’t possible to predict the distributive impact of increases in formality, it is feasible to establish a set of plausible hypothesis that point towards a relationship involving the employment share of each sector (formal/informal), the magnitude and dynamic of wage differentials and the inequality within each sector. Our empirical results based on a panel data set for Latin American countries suggest a positive association between wage inequality and the informality rate, which is mediated by the wage gap between the formal and informal sector, the intrasectoral wage inequality and the interaction term between informality and the wage gap. The results are robust to the inclusion of a set of usual variables used for the econometric identification of the determinants of inequality. They suggest that the process of formalization in the labor market, as the ones that took place in Latin American countries in the last decade, may be related to decreases in wage inequality, although the movements may differ depending on the magnitudes of the wages gaps and intrasectoral inequalities. Therefore, our results, along with the high persistence of informality in the region, point out
the need to understand in depth the relation between labor informality and wage inequality, a research area where advances are still scarce.

References


Schneider, F. (2012). The shadow economy and work in the shadow: what do we (not) know?. IZA Discussion Paper No. 6423

Annex

Composition effect

According to equation (1.8), the composition effect is:

\[
L^{1,x^2}(x^2, p) - L(p, x^1) = \left[ \frac{x^2 \mu^1_a}{\mu^{1,x^2}_a} L^1_a \left( \frac{F^1_a(z^{1,x^2})}{L^1_a(F^1_a(z^1))} \right) + \frac{(1 - x^2) \mu^1_b}{\mu^{1,x^2}_b} L^1_b \left( \frac{F^1_b(z^{1,x^2})}{L^1_b(F^1_b(z^1))} \right) \right] - \left[ \frac{x^1 \mu^1_a}{\mu^1} L^1_a \left( \frac{F^1_a(z^1)}{L^1_a(F^1_a(z^1))} \right) + \frac{(1 - x^1) \mu^1_b}{\mu^1} L^1_b \left( \frac{F^1_b(z^1)}{L^1_b(F^1_b(z^1))} \right) \right] (A.1)
\]

In order for the impact of an increase in formalization to be unambiguously equalizing, (A.1) must be equal or greater than zero for all \( p \). The expression could be rearranged as follows:

\[
\left[ \frac{x^2}{\mu^{1,x^2}_a} L^1_a \left( \frac{F^1_a(z^{1,x^2})}{L^1_a(F^1_a(z^1))} \right) - \frac{x^1}{\mu^1} L^1_a \left( \frac{F^1_a(z^1)}{L^1_a(F^1_a(z^1))} \right) \right] \mu^1_a + \left[ \frac{(1 - x^2)}{\mu^{1,x^2}_b} \left( \frac{F^1_b(z^{1,x^2})}{L^1_b(F^1_b(z^1))} \right) - \frac{(1 - x^1)}{\mu^1} \right] \mu^1_b \geq 0 (A.2)
\]

Moreover, taking out as a common factor \( L^1_a \left( \frac{F^1_a(z^1)}{L^1_a(F^1_a(z^1))} \right) \) and \( L^1_b \left( \frac{F^1_b(z^1)}{L^1_b(F^1_b(z^1))} \right) \) in each term, we obtain:

\[
\left[ \frac{x^2}{\mu^{1,x^2}_a} L^1_a \left( \frac{F^1_a(z^{1,x^2})}{L^1_a(F^1_a(z^1))} \right) - \frac{x^1}{\mu^1} \right] L^1_a \left( \frac{F^1_a(z^1)}{L^1_a(F^1_a(z^1))} \right) + \left[ \frac{(1 - x^2)}{\mu^{1,x^2}_b} \left( \frac{F^1_b(z^{1,x^2})}{L^1_b(F^1_b(z^1))} \right) - \frac{(1 - x^1)}{\mu^1} \right] \mu^1_b \mu^1_a L^1_a \left( \frac{F^1_a(z^1)}{L^1_a(F^1_a(z^1))} \right) \geq 0 (A.3)
\]

The latter condition can also be expressed as:

\[
\left[ \frac{x^2}{\mu^{1,x^2}_a} L^1_a \left( \frac{F^1_a(z^{1,x^2})}{L^1_a(F^1_a(z^1))} \right) - \frac{x^1}{\mu^1} \right] + \frac{\mu^1_b L^1_b \left( \frac{F^1_b(z^1)}{L^1_a(F^1_a(z^1))} \right)}{\mu^1_a L^1_a \left( \frac{F^1_a(z^1)}{L^1_a(F^1_a(z^1))} \right)} \geq 0 (A.4)
\]

Simplifying the notation, we can omit the Lorenz curve’s argument without loss of generality:

\[
L^1 \left( \frac{F^1(z^1)}{L^1_a(F^1_a(z^1))} \right) = L^1(z^1) \text{ and } L^1 \left( \frac{F^1(z^{1,x^2})}{L^1_a(F^1_a(z^1))} \right) = L^1 \left( \frac{F^1(z^{1,x^2})}{L^1_a(F^1_a(z^1))} \right).
\]

\[
\left[ \frac{x^2}{\mu^{1,x^2}} L^1_a \left( \frac{z^{1,x^2}}{L^1_a(z^1)} \right) - \frac{x^1}{\mu^1} \right] + \frac{\mu^1_b L^1_b \left( \frac{z^1}{L^1_a(z^1)} \right)}{\mu^1_a L^1_a \left( \frac{z^1}{L^1_a(z^1)} \right)} \geq 0 (A.5)
\]
In turn, naming $\beta$ the change in mean wages associated with the increase of formal sector employment share, it must be verified that $\mu^{1, x_2} = (1 + \beta)\mu^1$. Therefore, the previous expression can be expressed as follows:

$$\frac{x^2 \mu_1^2 L_a(z^{1, x_2^2}) - x^1 \mu_1^{1, x_2} L_a(z^{1})}{(1 - x^2) \mu_1^1 L_b(z^{1, x_2^2}) - (1 - x^1) \mu_1^{1, x_2} L_b(z^{1})} + \frac{\mu_1^1 L_b(z^{1})}{\mu_1^1 L_a(z^{1})} \geq 0 \quad (A.6)$$

Simplifying terms and considering that the Generalized Lorenz Curve is $\mu_i L_i(z)$,

$$\frac{x^2 L_a(z^{1, x_2^2}) - x^1 (1 + \beta) L_a(z^{1})}{(1 - x^2) L_b(z^{1, x_2^2}) - (1 + \beta)(1 - x^1) L_b(z^{1})} + \frac{L_G(z^{1})}{L_G(z^{1})} \geq 0 \quad (A.7)$$

Wage structure effect

$$L^{1, x_2} \mu^2 (x^2, p) - L^{1, x_2}(x^2, p)$$

$$= \left[ \frac{x^2 \mu_2^2}{\mu^2} L_a^1(z^{2}) + \frac{(1 - x^2) \mu_1^2}{\mu^2} L_b^1(z^{2}) \right]$$

$$- \left[ \frac{x^2 \mu_1^1}{\mu^1} L_a^1(z^{1}) + \frac{(1 - x^2) \mu_1^1}{\mu^1} L_b^1(z^{1}) \right] \quad (A.9)$$

Rearranging terms,

$$L^{1, x_2} \mu_1^2 (x^2, p) - L^{1, x_2}(x^2, p)$$

$$= x^2 \left[ \frac{\mu_2^2}{\mu^2} L_a^1(z^{2}) - \frac{\mu_1^1}{\mu^1} L_a^1(z^{1}) \right] + (1 - x^2) \left[ \frac{\mu_2^2}{\mu^2} L_b^1(z^{2}) - \frac{\mu_1^1}{\mu^1} L_b^1(z^{1}) \right] \quad (A.10)$$

Let $\alpha$ be the growth rate of mean wages in the overall labor market,

$$L^{1, x_2} \mu_1^2 (x^2, p) - L^{1, x_2}(x^2, p)$$

$$= \frac{x^2}{\mu^1} \left[ \frac{\mu_2^2}{(1 + \alpha)} L_a^1(z^{2}) - \frac{\mu_1^1}{(1 + \alpha)} L_a^1(z^{1}) \right]$$

$$+ \frac{(1 - x^2)}{\mu^1} \left[ \frac{\mu_2^2}{(1 + \alpha)} L_b^1(z^{2}) - \frac{\mu_1^1}{(1 + \alpha)} L_b^1(z^{1}) \right] \quad (A.11)$$

The condition for the effect of changes in the wage structure result in a reduction of inequality is $L^{1, x_2} \mu_1^2 (x^2, p) - L^{1, x_2}(x^2, p) \geq 0$, thus, and taking into consideration the definition of the Generalized Lorenz Curve, the wage structure effect will be equalizing if only:

$$x^2 \left[ \frac{L_G(z^{2})}{(1 + \alpha)} - L_G(z^{1}) \right] + (1 - x^2) \left[ \frac{L_G(z^{2})}{(1 + \alpha)} - L_G(z^{1}) \right] \geq 0 \quad (A.12)$$
Intra-sector inequality effect

\[
L^2(x^2, p) - L^{1,x^2, \mu^2_1}(x^2, p) \\
= \left[ \frac{x^2 \mu^2_a}{\mu^2} L^2_a(z^2) + \frac{(1 - x^2) \mu^2}{\mu^2} L^2_b(z^2) \right] \\
- \left[ \frac{x^2 \mu^2_a}{\mu^2} L^1_a(z^1) + \frac{(1 - x^2) \mu^2}{\mu^2} L^1_b(z^1) \right] \tag{A.13}
\]

\[
\frac{x^2 \mu^2_a}{\mu^2} [L^2_a(z^2) - L^1_a(z^1)] - \frac{(1 - x^2) \mu^2}{\mu^2} [L^2_b(z^2) - L^1_b(z^1)] \tag{A.14}
\]

Given that \( L^2(x^2, p) - L^{1,x^2, \mu^2_1}(x^2, p) \geq 0 \) must be met in order to ensure that this effect is equalizing, it must be fulfilled that:

\[
\frac{x^2 \mu^2_a}{(1 - x^2) \mu^2_b} \geq \frac{L^2_a(z^2) - L^1_a(z^1)}{L^2_a(z^2) - L^1_a(z^1)} \tag{A.15}
\]