ARTÍCULO

LABOUR INCOME INEQUALITY AND THE INFORMAL SECTOR IN COLOMBIAN CITIES

John Ariza
Gabriel Montes-Rojas

Ariza, J., & Montes-Rojas, G. (2017). Labour income inequality and the informal sector in Colombian cities. ***Cuadernos de Economía, 36*****(72), 77-98.**

Labour markets in developing countries are crucial to determine income inequality. In this paper, we use a panel data approach to study the effect of the informal sector on labour income inequality for thirteen cities in Colombia from 2002-2015. We use the rate of underemployment, the average duration of unemployment and the intensity of forced migration from armed conflicts as instruments for the urban informal sector. Results suggest that the informal sector has a positive and statistically significant effect on labour income inequality, which implies that an increase by one percentage point in the informal sector increases the Gini coefficient of labour income by about 0.07.

Keywords: Income inequality, the informal sector, Latin America, occupations.
JEL: J31, J46, J81, O54.

J. Ariza
PhD in Applied Economics. Universidad del Tolima. Departamento de Economía y Finanzas. Ibagué, Colombia. E-mail: jfariza@ut.edu.co.

G. Montes-Rojas
PhD in Economics. Universitat Autònoma de Barcelona. Departament d'Economia Aplicada. Barcelona, Spain. E-mail: gabriel.montes@uab.cat.

Sugerencia de citación: Ariza, J., & Montes-Rojas, G. (2017). Labour income inequality and the informal sector in Colombian cities. *Cuadernos de Economía, 36*(72), 77-98. doi: 10.15446/cuad.econ.v36n72.65820.

Este artículo fue recibido el 16 de agosto de 2016, ajustado el 13 de diciembre de 2016, y su publicación aprobada el 7 de febrero de 2017.

Ariza, J., & Montes-Rojas, G. (2017). Desigualdad de ingresos laborales y el sector informal en ciudades colombianas. *Cuadernos de Economía, 36*(72), 77-98.

Los mercados de trabajo en países en desarrollo son claves para determinar la desigualdad de ingresos. En este artículo usamos un enfoque de datos de panel para estudiar el efecto de la informalidad laboral sobre la desigualdad de los ingresos laborales para trece ciudades en Colombia durante el periodo 2002-2015. Instrumentamos la tasa urbana de informalidad laboral con la tasa de subempleo, la duración promedio del desempleo y con el flujo de personas de desplazamiento forzado. Los resultados sugieren que la informalidad laboral tiene un efecto positivo y estadísticamente significativo sobre la desigualdad de ingresos laborales. Un aumento de 1 punto porcentual en la tasa de informalidad laboral incrementa el coeficiente de Gini en 0,07.

Palabras clave: desigualdad de ingreso, informalidad laboral, América Latina, ocupaciones.
JEL: J31, J46, J81, O54.

Ariza, J., & Montes-Rojas, G. (2017). Inégalité des revenus du travail et le secteur informel dans les villes colombiennes. *Cuadernos de Economía, 36*(72), 77-98.

Les marchés de travail dans les pays en développement sont essentiels pour pour déterminer l'inégalité des revenus. Dans cet article, nous utilisons une approche de données de panel pour étudier l'effet de l'informalité du travail sur l'inégalité des revenus salariaux pour trois villes de Colombie pendant la période 2002-2015. Nous mettons en relation le taux urbain d'informalité du travail et le taux de sous-emploi, la durée moyenne du chômage et le flux de personnes du déplacement forcé. Les résultats suggèrent que l'informalité du travail a un effet positif et statistiquement significatif sur l'inégalité des revenus salariaux. Une augmentation de 1 pour cent du taux d'informalité du travail augmente le coefficient de Gini de 0,07.

Mots-clés : inégalité du revenu, informalité du travail, Amérique latine, occupations.
JEL: J31, J46, J81, O54.

Ariza, J., & Montes-Rojas, G. (2017). Desigualdade na receita dos trabalhadores e no setor informal nas cidades colombianas. *Cuadernos de Economía, 36*(72), 77-98.

Os mercados de trabalho nos países em desenvolvimento são peça chave para determinar a desigualdade de receitas. Nesse artigo usamos uma abordagem de dados em painel para estudar o efeito da informalidade trabalhista quanto à desigualdade nas receitas dos trabalhadores em treze cidades na Colômbia durante o período 2002-2015. Comparamos a estatística urbana de informalidade traba-

lhista com a de subemprego, a duração média do desemprego e levamos em conta o fluxo de retirantes forçados. Os resultados sugerem que a informalidade trabalhista tem um efeito positivo e estatisticamente significativo sobre a desigualdade na receita dos trabalhadores. Um aumento de 1 ponto percentual na estatística da informalidade trabalhista que incrementa coeficiente de Gini em 0,07.

Palavras chave: desigualdade de receita, informalidade trabalhista, América Latina, ocupações.
JEL: J31, J46, J81, O54.

INTRODUCTION

Income inequality in developing countries is an important issue in the research agenda and it has been subject to a great deal of public debate. Colombia and many other Latin American countries have high values of income inequality, rates that are only surpassed by African economies. Studies on the determinants of income concentration in the region suggest that the effect of economic growth, labour market forces, the strength of labour institutions, and fiscal policy are key factors to be able to explaining the recent decline in income inequality. Despite the region´s informal sector being large, not much attention has been paid to studying the implications that the size of this sector has on the wage dispersion.

Informal workers constitute an important share of the total amount of employment, and earnings from informal jobs are crucial in shaping the income distribution. Changes in the magnitude and in the composition of informal work determine distributional changes, mainly for the lower echelons of income distribution where most informal workers are located. In this paper, we study the relationship between the informal sector and income inequality. Considering that the specific characteristics of local labour markets play a central role in determining the quality of jobs and earnings, we use information from the main cities in Colombia during 2002-2015 to test the effect that the urban informal sector has on urban income inequality.

First, we study income inequality in cities by decomposing the Theil index into the between and within components from 2002 to 2015. This is a period characterised by having both high and moderate economic growth. We analyse the sources of variation of informal employment on a city level according to the definition of informality given by the Departamento Administrativo Nacional de Estadística (DANE). Second, we estimate the effect of the informal sector on income inequality by using a panel data approach to account for unobserved heterogeneity in urban areas. Since informality is only one of the determinants of income inequality, in the empirical strategy we also control for other factors such as education, industrial composition and the gender wage gap at city level.

Because we suspect that the level of informality is an endogenous variable, we propose three instrumental variables for the size of the informal sector in Colombian cities. Based on the fact that having negative expectations about finding a good formal job increases the probability of finding any job, we use the local level of underemployment, the duration of local unemployment, and the flow of forced migrants as instrumental variables to measure the level of informality. Higher values of these variables increase informality. Finally, we discuss the validity of our proposed instruments.

The remainder of the paper is organised as follows. In the next section we present a literature review of the link between informality and income inequality in developing countries. We then describe the empirical strategy and the data. The

next section presents styled facts and decomposing results. This is followed by the econometric results. Finally, we provide some concluding remarks.

LITERATURE REVIEW

There is a large quantity of literature that studies the determinants of income inequality and the informal sector in developing countries. The recent decline in income inequality in Latin America has been studied by Gasparini and Lustig (2011), Lustig, López-Calva and Ortiz-Juárez (2013), and Cornia (2014). These studies suggest that decreasing income inequality is a result of a combination of factors such as the fall in the returns from education, the increasing demand for low skilled workers (due to a commodities boom), and, in some countries, the strengthening of labour institutions and the implementation of social programmes such as monetary transfers.

In the same way as income inequality, the informal sector in Latin America increased during 1990s and then reduced during 2000s. After having studied the period between 1989-2005, Gasparini and Tornarolli (2009) suggest that the increase in informality in the 1990s was more consistent with the idea of voluntary self-employment than with the changes in the national employment structure that leaned towards the informal sectors. Maloney (2004) also wrote about this tendency. Moreover, the recent decrease in the informal sector in Brazil and Argentina has been documented in Berg (2010) and ILO (2011). They suggest that the decreasing pattern is associated with a better economic performance as well as with better labour institutions such as formalization and minimum wage programmes.

In terms of income inequality in Colombia, Arango, Posada and Uribe (2006) and Posso (2010) document the increase of income concentration during 1990s. They suggest that the increasing demand for high skilled workers is the main factor that is driving the wider wage differential between educated and less educated workers. To explain the recent decline in income inequality, Posso (2010) details that the decrease in returns from education justifies the changes that have occurred at the lower part of the income distribution. Studies about the informal sector in Colombia include Ortiz and Uribe (2004), and García (2009). These authors state that informality is related negatively to education and positively to cities' degree of industrial development as well as the size of firms. Galvis (2012) confirmed this idea in a recent study.

In terms of the specific relationship between the informal sector and income inequality in the region, three studies have recently analysed the matter. Maurizio (2014) considers the case of Argentina and Brazil during the 2000s. She stated that together with the decline in returns from education, the improvements in working conditions that were championed by certain labour institutions are relevant factors in explaining the decline of income inequality. The dynamic decomposition

of the Theil index in both countries suggests that the reduction in income inequality comes from the decrease of the within inequality component for both formal and informal workers.

Amarante and Arim (2015a) discuss the cases of Chile, Ecuador and Uruguay and also find a positive relationship between informality and income inequality. Despite the particularities of each country, they suggest that improvements in working conditions lead to not only welfare gains for workers but also lower levels of income concentration. By using a panel of sixteen Latin American countries and taking a timeframe of between 1990 and 2013, Amarante and Arim (2015b) estimate a positive and statistically significant informality coefficient for income inequality. However, they are aware that these results are not in terms of a cause - effect relationship.

More recently, Binelli (2016) contributed to the literature by studying the case of Mexico from 1987-2002. When considering salaried employees, she found that wage inequality for informal workers accounts for 60% of the total wage inequality. The co-movement between informality and inequality was studied econometrically by using the peso crisis of the mid-1990s as an instrument to determine the size of the informal sector. Results suggest that informality had a positive effect on wage inequality in different settings and for labour market institutions. She stated that higher dispersion among wage earners is one of the channels through which informality negatively affects development.

There are few studies that jointly analyse informality and income inequality on a city level in Colombia. The literature on Colombia mostly relates to the characteristics of the informal sector in cities while there are even less studies about income inequality in urban areas. The economic relevance of both social indicators has recently been drawn attention to in the form of public debate. Could national or local governments improve income disparities through implementing economic policies to reduce job informality? This paper contributes to this discussion by analysing the effect of the size of the informal sector on income distribution by using the main urban areas in the country as units of observation.

Our approach is empirically different from the Binelli (2016) work in the sense that we consider panel data methods and because we instrument job informality in cities using three variables: the level of underemployment, the average period of unemployment, and forced migration. Additionally, we believe that local labour market conditions, expressed by both individual dissatisfaction with the job and lower probability of being employed, induce the increase in the levels of informality. Therefore, bad expectations about local labour market performance could increase the level of informal jobs, which then affects the earnings mainly at the lowest part of the income distribution. Further analysis of the relationship between job informality and underemployment in Colombia can be found in Uribe, Ortiz and García (2008).

Furthermore, the local level of job informality could be greatly affected by the country's internal war. The rural armed conflict has resulted in huge flows of forced migration, and the displaced population end up in the main urban areas. This exogenous labour supply shock comprised of people with lower levels of education in cities puts pressure on the local labour markets and increases the chances that people will undertake any economic activity in order to obtain income. As a result, the level of informal work in cities has risen. A recent study on this topic can be found in Rozo and Winkler (2016). They found that the large inflows of displaced persons induced a sizable negative effect on the performance of formal firms.

In our approach, we consider the following to be relevant factors that determine income inequality on a city level: the effect of the education, the industrial structure and the gender wage gap. For the first factor, we take into account the proportion of workers with a tertiary education since in cities the increase in average years of schooling could *ceteris paribus* raise the wage gap between more and less educated people: more education implies higher wages. For the second factor, the composition of economic activity in cities determines labour productivity and wages. The manufacturing sector having a greater share of the economy in urban areas is correlated with higher wages and also with a wider difference in wages between workers in the manufacturing sector and workers in the other sectors. Another source of income inequality on a city level is gender discrimination in the labour market. We use the gender wage gap as a measure of this phenomenon. We use the Mincerian wage equations to obtain the coefficients of the gender wage differentials. The widening wage gap between men and women with the same observable characteristics increases the level of income inequality in cities.

EMPIRICAL STRATEGY

In order to estimate the effect of informality on income inequality, we use panel data methods. The panel data approach provides a general framework that explicitly models unobserved heterogeneity among individuals. Since we have a sample of Colombian cities that has been observed over time, we could econometrically study the effect once we control for unobserved heterogeneity on a city level. Following Wooldridge (2010), the panel data model can be presented as

$$y_{it} = \beta_0 + x_{it}\beta + c_i + u_{it} \tag{1}$$

where y_{it} refers to the dependent variable for individual i in time period t, β_0 is the intercept of the equation that is common to all individuals, x is a matrix of observable variables that affects y; β is a vector of slopes, c_i is an individual unobservable random variable (individual heterogeneity), and u_{it} is the idiosyncratic error that changes across i and t. For our purpose, equation (1) becomes

$$Gini_{it} = \beta_0 + \beta_1 Inf_{it} + \beta_2 Te_{it} + \beta_3 M_{it} + \beta_4 Gwg_{it} + c_i + u_{it} \quad (2)$$

where the dependent variable is the Gini coefficient of income in city *i* during time period *t*, *Inf* is the rate of informality, *Te* is the percentage of the working population who have completed their tertiary education, *M* is the proportion of people employed in the manufacturing sector, and *Gwg* is the city gender wage gap. Terms c_i and u_{it} are the same as above. We implemented the above model with fixed-effects (FE) and random-effects (RE) estimators. For the latter, we checked for its validity using the Hausman (1978) test.

Controlling for the unobserved heterogeneity when accounting for the causal effect in regression is not enough to be able to control for other potential sources of endogeneity. In this model, explanatory variables may not be exogenous, and reverse causality could be a potential source of endogeneity. The informal sector is the result of local labour market conditions, which, in terms of our model, implies that it is related to the error term, *i.e*. unobserved causes of inequality. In order to solve this problem, we considered an instrumental variables (IV) approach. In order to use this, we proposed three instruments that explain the level of informal employment.

The first instrument is the rate of underemployment. This measure shows the level of worker dissatisfaction with their own job and indicates the limited possibilities to increase productivity and income in their actual jobs. The level of underemployment could explain workers´ reallocation from lower-profit to higher-profit sectors, and, thus, it is related to levels of informality in cities. It could be argued that we cannot rule out the possibility that underemployment is itself related to inequality levels. This could, however, be the result of other latent structural factors for which we control using the additional covariates.

The second instrument is the duration of unemployment. Longer time periods searching for a job affect people negatively and the result is that they end up taking any job in order for them to achieve a minimum level of income. This implies that labour market conditions affect the quality of employment in terms of stability and salary. Therefore, long periods of unemployment are related to higher levels of labour informality.

The third instrument is forced migration. Colombia has suffered an internal war for many years that had resulted in huge levels of forced migration throughout the whole country. People from rural areas migrate to the cities to flee the war and find new sources of income. As a result, people arriving in cities put pressure on labour market indicators. Usually, these people have low levels of education and end up working in informal jobs. For this indicator, we have used the intensity of the population inflow as a result of forced migration. In the model, forced migration is a dummy variable that indicates if the city receives more than one thousand displaced people in the current year.

Although we cannot rule out a potential correlation between the instruments and the error term in the panel data equation, we can assume that these have a much stronger correlation with informality than with inequality. If indeed there is a correlation between the IV and the error term, this would be positive.

Thus, we estimate equation (2) by IV, using the underemployment rate, the unemployment duration, and the forced migration as instruments. The validity of the instruments is tested by the Sargan test of overidentifying restrictions where the null hypothesis states that all excluded instruments are valid instruments. The validity of the instruments means that instruments are strictly exogenous with respect to the idiosyncratic error (depending if it is FE or RE), and they can either be correlated or not with the unobservable heterogeneity component.

DATA

The data used in this paper comes from the national household survey carried out by the DANE. The surveys collect information about sociodemographic characteristics and labour market indicators. We use two household surveys from two different time periods as one was an updated version of the other. The particular surveys used were the *Encuesta Continua de Hogares* (ECH) 2002-2006 and the *Gran Encuesta Integrada de Hogares* (GEIH) 2007-2015. Although the latter survey included more cities, for the sake of consistency we used the same thirteen cities and metropolitan areas that were surveyed in the ECH from 2002. We use the second quarter of each year because it was the only period in which informality was measured in the ECH. We do not use micro-data before the 2000s as the household surveys in that time period are not comparable with the ECH and the GEIH.

We built a panel of 13 cities and metropolitan areas, taking the data from the second quarter of each year from 2002-2015. The data contained the following indicators: income inequality, labour informality, and the other variables considered in the empirical strategy. To estimate the income inequality measurements, we took different types of labour income and then we deflated them by using the Consumer Price Index given by the DANE. Specifically, we calculated both the Theil Index and the Gini coefficient for wages and for the total labour income; this included income from wage earners, self-employed workers, domestic servants and employers. We did not impute any value from wages, earnings or other labour incomes, and, as such, we ruled out this type of missing data from the sample. The forced migration data on a city level were obtained from official statistics from the Colombian government's *Unidad de Víctimas*.

To measure job informality, we first followed the approach recommended by DANE that stated using seven items regarding firm size and job characteristics. According to DANE, a small firm has up to five employees. The definition of an informal worker includes employees in any of the following situations:

- Private wage earners who work in small firms
- Non-professionals who are self-employed and work in small firms
- Unpaid work undertaken by family members in small firms
- Unpaid workers in firms
- Domestic servants working for small firms
- Labourers working for small firms
- Employers in small firms

We compute the aforementioned indicators on a city level in order to study the specific components of labour informality. However, and because this is a particular measure of the Colombian informal labour, in our empirical work, we consider another measure based on the "legal" definition that is broadly used on an international level: the percentage of workers who are not affiliated to either health or pension schemes.

RESULTS

Labour income inequality in Colombian cities

In Table 1 we present two measures of labour income inequality for all types of workers. Columns (1) and (2) report the Gini coefficient and the Theil Index for total labour income, according to which, at the beginning of the 2000s the labour income inequality for the thirteen main cities was around 0.50 for the first measure and 0.53 for the second. When analysing the Theil index, the two main cities at the top of the most unequal areas in Colombia are Montería and Bogotá; these are the smallest and the biggest cities in the sample, respectively. Thus, it can be said that there is no clear pattern between city size and income concentration.

In contrast, the drop in the labour income inequality reported on a national level was a general pattern in all urban areas. For almost all cities, both the Gini coefficient and the Theil index decreased between 2002-2015 (columns 3 and 4). Since these two measurements are sensitive to changes at different parts of the income distribution (the Gini to changes in the centre and the Theil to changes at the upper tail), they provide evidence that the decrease was pervasive. The highest drop took place in cities such as Pereira, Manizales and Bogotá. These high drops for the most unequal cities can only be good news.

Table 1.
Labour income inequality in Colombia

	2002		2002-2015	
	Gini index	**Theil index**	**Δ Gini index**	**Δ Theil index**
	(1)	**(2)**	**(3)**	**(4)**
Thirteen areas:	0.50	0.53	-0.05	-0.10
City				
Bogotá	0.53	0.55	-0.07	-0.12
Medellín	0.50	0.49	-0.02	-0.01
Cali	0.50	0.51	-0.06	-0.06
Barranquilla	0.50	0.48	-0.06	-0.11
Bucaramanga	0.46	0.42	-0.04	-0.08
Cartagena	0.43	0.34	0.00	-0.01
Cúcuta	0.45	0.40	-0.03	-0.06
Pereira	0.50	0.59	-0.06	-0.14
Ibagué	0.51	0.48	-0.04	-0.08
Villavicencio	0.48	0.43	-0.06	-0.03
Manizales	0.51	0.49	-0.08	-0.12
Pasto	0.53	0.53	-0.01	0.08
Montería	0.56	0.66	-0.14	-0.32

Source: Author's calculations based on micro data from ECH and GEIH. All calculations use sample weights.

In Table 2 we present the inequality measurements by type of employment. According to the results, labour income is highly concentrated in the self-employed and employers groups. The wage earners group is the most important in terms of population share and income share; it represents about 55 percent of the total employment and 64 percent of the total labour income. So, the distributional changes in income for this group constitute one of the key factors to be able to disentangle aggregate changes in labour income inequality. During this period, the self-employed group increased its share in terms of both total employment and total income.

Table 3 shows the results from the decomposition of the Theil index into the between and within components. Unlike the Gini coefficient, the Theil index belongs to the Generalized Entropy index family, which is additively decomposable. According to the results, the within component accounts for more than 80% of the income inequality both in terms of cities and employment status. If we look at the changes, the drop in the cities' income inequality was mainly explained by the drop in the within component while for employment position this fall was equally distributed.

Table 2.
Labour income inequality by type of employment

	2002				2002-2015			
	Gini Index	**Theil index**	**Pop. share**	**Income share**	**Δ Gini Index**	**Δ Theil index**	**Δ Pop. share**	**Δ Income share**
Wage earner	0.44	0.40	0.55	0.64	-0.05	-0.06	-0.01	-0.00
Self-employed	0.55	0.60	0.32	0.22	-0.03	-0.06	-0.03	-0.05
Domestic Servant	0.29	0.15	0.08	0.03	-0.01	-0.00	-0.04	-0.01
Employers	0.53	0.53	0.05	0.11	-0.06	-0.08	-0.01	-0.04
Other	0.39	0.25	0.00	0.00	-0.04	-0.05	-0.00	-0.00

Source: Author's calculations based on micro data from ECH and GEIH. Results were obtained using the *ineqdeco* programme in Stata. All calculations use sample weights.

The role of the within component for cities is other way to report the decreasing pattern in the income inequality of urban areas that was previously mentioned. Regarding the high participation of this within component in 2002 and its change over time, results would suggest no differentiated effects of inequality trends among cities. This, in turn, could be related to the lack (or the non-significant effect) of local policies aimed at reducing income inequality.

Table 3.
Theil index decomposition

	For cities					For employment position				
	2002		**2015**		**Change**	**2002**		**2015**		**Change**
Theil	0.534		0.438		-0.10	0.534		0.438		-0.10
Between	0.016	3%	0.007	2%	-0.01	0.081	15%	0.035	8%	-0.05
Within	0.519	97%	0.431	98%	-0.09	0.454	85%	0.404	92%	-0.05

Source: Author's calculations based on micro data from ECH and GEIH. Results were obtained using the *ineqdeco* programme in Stata. All calculations use sample weights.

Job informality in Colombian cities

Table 4 displays the components of the job informality measurement, which is consistent with the DANE's definition. Columns (1) and (2) report the initial informality levels and their changes during the period of analysis. According to the table, 67% of the workers in the main cities had an informal job in 2002. Cúcuta, Montería and Villavicencio, cities with no more than 700,000 inhabitants, reported

rates of informality that were higher than 73%. The largest cities, Bogotá and Medellín, by contrast, had informality rates below 60%. Throughout the period of analysis, job informality considerably decreased in all cities.

Columns (3) to (9) show the distribution of informal workers and its variation over time on both an aggregate and city level. Accordingly, by 2002 the self-employed non-professionals represented almost half of the total amount of informality while wage earners working in small firms represented about 28%. It is clear that from 2002-2015 informality decreased among wage earners and domestic servants while it increased for the self-employed non-professionals. This pattern remains constant for all cities. Bogotá contributed with 38% of total increase in the informal self-employed: 12.4 percentage points.

Econometric results

Table 5 presents average and standard deviations of the variables for the thirteen cities considered in the econometric model. The Gini coefficient is higher for the more heterogeneous groups, and labour informality is greater for the DANE measurement. During the period, labour market indicators reported more skilled workers, higher levels of gender discrimination among wage earners, and a lower proportion of workers in the manufacturing sector. Regarding the instruments considered, the three variables also report decreases between 2002-2015. Forced migration, for example, went from 0.25% of total city inhabitants in 2002 to 0.02% in 2015.

In Table 6 we present the econometric results of the effect of job informality on the Gini coefficient for different kinds of labour incomes. The income inequality measurements were constructed by the following groups: wage earners, wage earners plus the self-employed, and for all workers (including domestic service and employers). Columns 1 and 2 show estimations for the fixed-effects and random-effects (RE) models, respectively for the first group. The results indicate that the effect of job informality on wage concentration on a city level is positive and statistically significant for both specifications.

The Hausman test suggests that the RE estimator is valid (column 2). This implies that a decrease of one percentage point in job informality reduces the Gini coefficient among wage earners by about 0.0016. However, in line with the previous discussion, these estimations could be biased since the level of informality is an endogenous variable. Results for the instrumental variables are presented in a RE model, which is presented in columns 3 and 4, respectively. In this model, the effect is positive, statistically significant and stronger than the previous one. The effect of informality on inequality is almost three times higher than the informality level that was obtained without instruments. Once we employ the instrumental variables, the positive effect is consistent with the results contained in the literature (Amarante & Arim, 2015b; Binelli, 2016).

Table 4.
Job informality in Colombia

	Informality rate		Distribution of job informality						
	2002	Change 2002-2015	1. Wage earners in small firms	2. Self-employed non- professionals	3. Unpaid work by family members	4. Unpaid workers in firms	5. Domestic servants	6. Labourers	7. Employers of small firms
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Thirteen areas	67.2	-13.6	28.6	47.4	4.9	0.2	10.0	0.0	9.0
Change 2002-2015			-7.63	12.45	-0.58	0.31	-2.19	0.07	-2.44
City									
Bogotá	55.8	-11.8	-2.23	4.84	-0.55	0.09	-0.22	0.01	-0.65
Medellín	59.0	-15.5	-0.32	0.26	0.03	0.01	-0.48	0.01	-0.44
Cali	62.7	-16.0	-1.72	0.76	-0.18	0.08	-0.48	0.00	-0.48
Barranquilla	64.0	-8.9	0.02	1.47	0.14	0.03	-0.10	0.00	0.01
Bucaramanga	66.2	-11.7	-0.74	1.30	-0.01	0.00	-0.13	0.01	-0.06
Cartagena	68.5	-11.8	-0.25	0.75	0.13	0.02	-0.07	0.00	0.00
Cúcuta	79.1	-8.9	-1.19	1.56	-0.03	0.02	-0.25	0.00	-0.28
Pereira	64.6	-14.7	-0.38	0.18	0.02	0.01	-0.17	0.01	-0.13
Ibagué	70.2	-11.8	-0.12	0.23	-0.06	0.03	-0.10	0.01	-0.05
Villavicencio	73.8	-16.3	-0.06	0.44	-0.07	0.00	-0.06	0.00	-0.08
Manizales	62.1	-21.1	-0.32	-0.13	-0.04	0.01	-0.05	0.01	-0.05
Pasto	72.8	-14.1	-0.19	0.48	0.00	0.01	-0.05	0.00	-0.12
Montería	74.4	-14.0	-0.13	0.29	0.04	0.00	-0.02	0.00	-0.10
Total			-7.63	12.45	-0.58	0.31	-2.19	0.07	-2.44

Source: Author's calculations based on micro data from ECH and GEIH. All calculations use sample weights.

Table 5.
Variables in the econometric estimation

	2002		2015	
Variables	**Mean**	**Sd.**	**Mean**	**Sd.**
Labour income inequality				
Gini coefficient for wage earners	0.42	0.04	0.36	0.04
Gini coefficient for wage earners + self-employed	0.47	0.04	0.43	0.03
Gini coefficient for all employees	0.50	0.04	0.44	0.03
Labour informality				
Informality (DANE)	0.62	0.07	0.48	0.08
Informality (Health affiliation)	0.44	0.10	0.38	0.11
Other explanatory variables				
Proportion of high skilled workers	0.13	0.03	0.15	0.02
Proportion of workers in manufacturing	0.17	0.04	0.14	0.04
Gender wage gap	0.03	0.04	0.08	0.03
Instruments				
Proportion of workers in underemployment	0.33	0.06	0.26	0.08
Unemployment duration in months	4.25	0.66	2.77	0.69
Proportion of forced migration	0.02	0.01	0.002	0.00

Source: Author's calculations based on micro data from ECH and GEIH. By definition, the simple mean values of variables differ from the aggregate value of thirteen cities in Colombia that were reported by the DANE in its official statistics.

In terms of both the relevance and the exogeneity condition of the instruments, we present the Sargan test in the lower part of the Table 6 and the first stage estimations in Table 8. According to the value of the Sargan test, the set of instruments in the model are valid. In general, the three instruments proposed meet the two conditions among wage earners. For the other two groups of workers, we present RE estimations that also exclude and include forced migration in the set of instruments in columns 5 to 8. The statistical significance of the coefficients and the validity conditions for the instruments remain constant. The effect of job informality on labour income inequality increases when a more heterogeneous group of workers is considered.

In Table 7 we present the econometric results for the other measure of job informality, *i.e.* the proportion of workers without health and pension affiliation. Similarly to the previous case, we present the results for the three groups of workers, the respective test for exogeneity as well as the test to verify the instruments. The effect of job informality on income inequality, using a FE model, remains positive and statistically significant. Unlike previous results, the inclusion of forced migration

Table 6.
Effect of job informality on the labour income Gini coefficient

		Informality as defined by the DANE							
		Wage earners							
		Fixed-Effects		Random-Effects		REIV_1		REIV_2	
		(1)		(2)		(3)		(4)	
Informality		0.153	**	0.157	***	0.449	***	0.487	***
		(0.062)		(0.053)		(0.115)		(0.107)	
Highly skilled		0.105		0.301	*	0.610	***	0.655	***
		(0.177)		(0.165)		(0.204)		(0.203)	
Manufacturing		0.506	*	0.441	***	0.308	**	0.295	*
		(0.186)		(0.138)		(0.154		(0.154)	
Gender wage gap		0.066		0.046		0.169	*	0.184	**
		(0.076)		(0.075)		(0.092)		(0.092)	
Constant		0.207	***	0.189	***	-0.014		-0.042	
		(0.052)		(0.051)		(0.088)		(0.084)	
Underemployment						Yes		Yes	
Unemployment duration						Yes		Yes	
Forced migration						No		Yes	
N		182		182		182		182	
Sigma_u		0.04		0.03		0.03		0.03	
Sigma_e		0.03		0.03		0.04		0.04	
Rho		0.56		0.41		0.39		0.39	
Hausman test	Chi2 (4)			9.28					
	Prob>Chi2			0.055					
Sargan test	Chi2(df)					1.64		2.19	
	P-value					0.20		0.34	

Table 6.
Effect of job informality on the labour income Gini coefficient (*continued*)

		Informality as defined by the DANE							
		Wage earners + Self-employed				**All workers**			
		REIV_3		**REIV_4**		**REIV_5**		**REIV_6**	
		(5)		**(6)**		**(7)**		**(8)**	
Informality		0.498	***	0.502	***	0.554	***	0.533	***
		(0.112)		(0.103)		(0.115)		(0.104)	
Highly skilled		0.592	***	0.603	***	0.637	***	0.616	***
		(0.199)		(0.194)		(0.205)		(0.197)	
Manufacturing		-0.010		-0.003		0.080		0.090	
		(0.151)		(0.149)		(0.156)		(0.152)	
Gender wage gap		0.136		0.138		0.138		0.130	
		(0.089)		(0.087)		(0.092)		(0.089)	
Constant		0.066		0.061		0.034		0.049	
		(0.085)		(0.080)		(0.088)		(0.081)	
Underemployment		Yes		Yes		Yes		Yes	
Unemployment duration		Yes		Yes		Yes		Yes	
Forced migration		No		Yes		No		Yes	
N		182		182		182		182	
Sigma_u		0.03		0.03		0.03		0.03	
Sigma_e		0.04		0.04		0.04		0.036	
Rho		0.41		0.40		0.41		0.406	
Hausman test	Chi2 (4)								
	Prob>Chi2								
Sargan test	Chi2(df)	2.26		2.22		3.03		3.33	
	P-value	0.13		0.33		0.08		0.19	

Notes: *** $p < 0.01$; ** $p < 0.5$; * $p < 0.1$. Endogenous variable: informality rate. Instruments: highly skilled, manufacturing, gender wage gap, underemployment rate, duration of unemployment, and forced migration. Standard errors in brackets.

Table 7.
Effect of job informality on the labour income Gini coefficient

		Informality defined by health and pension affiliation							
		Wage earners							
		Fixed-Effects		**Random-Effects**		**FEIV_1**		**FEIV_2**	
		(1)		**(2)**		**(3)**		**(4)**	
Informality		0.379	***	0.194	***	2.018	*	1.290	***
		(0.104)		(0.065)		(1.077)		(0.427)	
Highly skilled		0.305		0.332	**	1.670	*	1.064	*
		(0.189)		(0.167)		(0.936)		(0.409)	
Manufacturing		0.474	***	0.535	***	-0.537		-0.088	
		(0.176)		(0.134)		(0.713)		(0.329)	
Gender wage gap		0.060		0.016		0.261		0.171	*
		(0.073)		(0.073)		(0.174)		(0.101)	
Constant		0.120	*	0.185	***	-0.597		-0.279	
		(0.060)		(0.051)		(0.475)		(0.193)	
Underemployment						Yes		Yes	
Unemployment duration						Yes		Yes	
Forced migration						No		Yes	
N		182		182		182		182	
Sigma_u		0.05		0.03		0.23		0.14	
Sigma_e		0.03		0.03		0.05		0.04	
Rho		0.69		0.40		0.95		0.92	
Hausman test	Chi2 (4)			16.72					
	Prob>Chi2			0.002					
Sargan test	Chi2(df)					1.79		4.14	
	P-value					0.18		0.13	

Table 7.
Effect of job informality on the labour income Gini coefficient (*continued*)

		Informality defined by health and pension affiliation							
		Wage earners + Self- employed				**All workers**			
		FEIV_3		**FEIV_4**		**FEIV_5**		**FEIV_5**	
		(5)		**(6)**		**(7)**		**(8)**	
Informality		2.614	*	1.374	***	2.883	*	1.286	***
		(1.325)		(0.422)		(1.454)		(0.417)	
Highly skilled		2.128	*	1.095	***	2.350	*	1.019	**
		(1.152)		(0.405)		(1.263)		(0.400)	
Manufacturing		-1.341		-0.576	*	-1.349		-0.363	
		(0.877)		(0.326)		(0.962)		(0.322)	
Gender wage gap		0.261		0.109		0.279		0.083	
		(0.214)		(0.100)		(0.235)		(0.099)	
Constant		-0.724		-0.181		-0.843		-0.145	
		(0.585)		(0.191)		(0.641)		(0.188)	
Underemployment		Yes		Yes		Yes		Yes	
Unemployment duration		Yes		Yes		Yes		Yes	
Forced migration		No		Yes		No		Yes	
N		182		182		182		182	
Sigma_u		0.30		0.16		0.33		0.15	
Sigma_e		0.07		0.04		0.07		0.04	
Rho		0.95		0.94		0.95		0.93	
Hausman test	Chi2 (4)								
	Prob>Chi2								
Sargan test	Chi2(df)	1.83		8.06		2.08		12.28	
	P-value	0.18		0.02		0.15		0.00	

Notes: *** $p < 0.01$; ** $p < 0.5$; * $p < 0.1$. Endogenous variable: informality rate. Instruments: highly skilled, manufacturing, gender wage gap, underemployment rate, duration of unemployment, and forced migration. Standard errors in brackets.

Table 8.
First-stage regression. REIV estimations

	Informality by DANE						**Informality by health and pension affiliation**					
	Wage earners		**Wage earners + Self-employed**		**All workers**		**Wage earners**		**Wage earners + Self-employed**		**All workers**	
	REIV_2		**REIV_4**		**REIV_6**		**FEIV_2**		**FEIV_4**		**FEIV_5**	
Highly skilled	-0.77	***	-0.76	***	-0.75	***	-0.73	***	-0.82	***	-0.82	***
	(0.19)		(0.19)		(0.19)		(0.12)		(0.13)		(0.13)	
Manufacturing	-0.18		-0.17		-0.15		0.36	**	0.50	***	0.50	***
	(0.18)		(0.18)		(0.18)		(0.14)		(0.13)		(0.13)	
Gender wage gap	-0.28	***	-0.28	***	-0.28	***	-0.12	*	-0.11	*	-0.11	*
	(0.09)		(0.09)		(0.09)		(0.05)		(0.05)		(0.05)	
Underemployment	0.09	*	0.09	*	0.09	*	-0.01		0.00		0.00	
	(0.05)		(0.05)		(0.05)		(0.03)		(0.03)		(0.03)	
Unemployment duration	0.03	***	0.03	***	0.03	***	0.01	**	0.01	*	0.01	*
	(0.01)		(0.01)		(0.01)		(0.00)		(0.00)		(0.00)	
Forced migration	0.05	***	0.05	***	0.05	***	0.03	***				
	(0.01)		(0.01)		(0.01)		(0.01)					
Constant	0.56	***	0.56	***	0.55	***	0.42	***	0.43	***	0.43	***
	(0.05)		(0.05)		(0.05)		(0.03)		(0.03)		(0.03)	

Notes: *** $p < 0.01$; ** $p < 0.5$; * $p < 0.1$. Endogeneous variable: informality rate. Instruments: highly skilled, manufacturing, gender wage gap, underemployment rate, duration of unemployment, and forced migration. Standard errors in brackets.

to the set of instruments is only relevant for the wage earners group. This could be due to the measure of informality used being too broad. Despite this, the effect is higher for the more heterogeneous group of workers when we make the estimation using instrumental variables.

In summary, the estimation of the effect of labour informality on income inequality is positive, statistically significant for both definitions of informality, and increases when a more heterogeneous group of workers is used. According to the econometric results and the trends in labour informality, the average aggregate labour informality decrease of thirteen percentage points that was reported by DANE during the period would have reduced the Gini coefficient among wage earners by about

0.06 and by about 0.07 for all workers. When we consider health and pension affiliation, the six percentage points decrease in informality would have reduced the Gini coefficient for wage earners by about 0.07 and by about 0.17 for all workers.

CONCLUSIONS

The decline in labour income inequality that has taken place in Colombia since 2002 is a growing topic of research. However, despite the potential distributional effects that changes in informality may have on labour income distribution, there have been only a few papers that have addressed this issue. Estimating the effect of job informality on income inequality allows us to contribute to the discussion regarding to what extent national or local governments can improve income disparities through implementing policies that reduce labour informality.

By considering a panel data of Colombian cities from 2002 to 2015, we have discussed and presented an estimation of the causal effect that job informality has had on the labour income Gini coefficient. Based in the premise that the specific labour market conditions in Colombian cities determine earnings and levels of labour income inequalities, we have found that after controlling for endogeneity in informal jobs, a one percentage point decrease in the informality rate reduces the Gini coefficient for wage earners by about 0.06 and by about 0.07 for all workers.

REFERENCES

1. Amarante, V., & Arim, R. (2015a). *Desigualdad e informalidad. Un análisis de cinco experiencias latinoamericanas*. Santiago de Chile: CEPAL, Ministerio de Asuntos Exteriores de Noruega.
2. Amarante, V., & Arim, R. (2015b). Desigualdad de las remuneraciones e informalidad: breve revisión de literatura y marco analítico. In V. Amarante, & R. Arim (Eds.), *Desigualdad e informalidad. Un análisis de cinco experiencias latinoamericanas* (pp. 19-35). Santiago de Chile: CEPAL, Ministerio de Asuntos Exteriores de Noruega.
3. Arango, L., Posada, C., & Uribe, J. (2006). Cambios en la estructura de los salarios urbanos en Colombia, 1984-2000. *Lecturas de Economía*, *64*, 187-194.
4. Berg, J. (2010). *Laws or luck? Understanding rising formality in Brazil in the 2000s*. Munich Personal RPEC Archive MPRA.
5. Binelli, C. (2016). Wage inequality and informality: Evidence from México. *IZA Journal of Labor & Development, 5*(5), 1-18. doi 10.1186/s40175-016-0050-1.

6. Cornia, G. (2014). Inequality trends and their determinants: Latin America over the period 1990-2010. In G. Cornia (Ed.) *Falling inequality in Latin America* (pp. 23-48). New York: Oxford University Press.
7. Galvis, L. (2012). Informalidad laboral en las áreas urbanas en Colombia. *Coyuntura Económica: Investigación Económica y Social, 42*(1), 15-51.
8. García, G. (2009). *Evolución de la informalidad laboral en Colombia: determinantes macro y efectos locales* (Archivos de Economía. Documento 360). Bogotá: Departamento Nacional de Planeación.
9. Gasparini, L., & Tornarolli, L. (2009). Labor informality in Latin America and the Caribbean: Patterns and trends from household survey microdata. *Desarrollo y Sociedad*, *63*, 13-80.
10. Gasparini, L., & Lustig, N. (2011). The rise and fall of income inequality in Latin American. In J. Ocampo & J. Ros (Eds.), *The Oxford Handbook of Latin American Economics* (pp. 691-714). New York: Oxford Handbooks in Economics.
11. Hausman, J. (1978). Specifications test in econometrics. *Econometrica*, *46*, 1251-1271.
12. International Labour Organization [ILO] (2011). *La formalización del empleo en Argentina.* Lima: Notas OIT. Trabajo decente en Argentina.
13. Lustig, N., López-Calva, E., & Ortiz-Juárez, E. (2013). Declining inequality in Latin America in the 2000s: The cases of Argentina, Brazil, and México. *World Development*, *44*, 129-141. doi: 10.1016/j.worlddev.2012.09.013.
14. Maloney, W. (2004). Informality revisited. *World Development*, *32*(7), 1159-1178. doi: 10.1016/j.worlddev.2004.01.008.
15. Maurizio, R. (2014). *Labour formalization and declining inequality in Argentina and Brazil in 2000s* (Research Paper 9). ILO.
16. Ortiz, C., & Uribe, J. (2004). *Características de la informalidad urbana en las diez principales áreas metropolitanas de Colombia: 1988-2000* (Documentos de Trabajo CIDSE). Universidad del Valle.
17. Posso, C. (2010). Desigualdad salarial en Colombia 1984-2005: cambios en la composición del mercado laboral y retornos a la educación postsecundaria. *Desarrollo y Sociedad*, *63*, 65-113.
18. Rozo, S., & Winkler, H. (2016). *Are immigration flows good for business? The role of informality in developing countries* (Working Paper). SSRN https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2819225.
19. Uribe, J., Ortiz, C., & García, G. (2008). Informalidad y subempleo en Colombia: dos caras de la misma moneda. *Cuadernos de Administración*, *21*(37), 211-241.
20. Wooldridge, J. (2010). *Econometric analysis of cross section and panel data* (2nd Ed.). Cambridge, United States of America: The MIT Press.