CRISTINA FERNÁNDEZ
Fedesarrollo

LEONARDO VILLAR
Fedesarrollo

The Impact of Lowering the Payroll Tax on Informality in Colombia

ABSTRACT In 2012, the Colombian Government reduced employer payroll contributions from 29.5 to 16.0 percent. Two years later, the informality rate had diminished by about 4.0 percentage points. This paper attempts to estimate how much of this reduction was due to the tax reform, isolating the impact of other macroeconomic variables. A natural approach to performing this task is to apply a difference-in-differences methodology using a household survey panel. Since the Colombian survey does not have a panel structure, we simulated one using a matching difference-in-differences methodology. According to the results, the tax reform is associated with a 4.8-percentage-point decrease in the informality of workers affected by the reform in the thirteen main metropolitan areas. This represents approximately half the reduction of the relevant informality rate during that period, affecting mostly salaried men and workers in general with low levels of education.

JEL Codes: J460, H230, C210

Keywords: Informal markets, payroll taxes, matching, difference-in-differences

In 2012, the Colombian government reformed the tax law by reducing payroll contributions from 29.5 to 16.0 percent of the monthly wage and substituting them with a profit tax.[1] The reform only affected payments made by employers of two or more workers that earn wages between one and ten times the minimum wage, and it did not change the amount of taxes or contributions payable by the workers. Nongovernmental organizations, the government, single-person businesses, and self-employed individuals were excluded from the reform. From a fiscal perspective, the source of these revenues was shifted

ACKNOWLEDGMENTS The authors want to thank Paulo Sánchez, Nicolás Gómez, and Francisco Fernández for their excellent research assistance; Anil Sinwal and Juan Villa for their technical support; and Mariano Bosch, Marcela Eslava, Carlos Medina, and the macroeconomic analysis and research groups of the CAF-Development Bank of Latin America, the Institute of Development Studies (IDS) of the University of Sussex, and the Bank of the Republic of Colombia for their valuable comments. This project was supported by Practical Action (PAC-DFID) and CAF.

1. Law 1607 of 2012.

to a profit tax (CREE) under the hypothesis that it is preferable to tax capital than to tax labor.[2]

In December 2014, two years after the law was passed, labor-informality rates in the thirteen main Colombian metropolitan areas had dropped from 56 to 52 percent.[3] When smaller cities and rural areas are included, the reduction was from 68 to 64 percent. These results hold using different measures of informality. The period after the reform was also characterized by high yet diminishing growth rates, changes in the tax rates, and increasing real minimum wages. What we are mainly interested in is knowing how much of the reduction in the informality rate was due to the tax reform.

Two empirical facts support the hypothesis that part of the recent reduction of informality in Colombia was due to the tax reform and not only to growth. First, the relationship between growth and informality weakened after the reform. The coefficient of correlation between the output gap and informality was –0.9 between 2001 and 2012 and –0.7 for the 2001–2015 period, signaling that something other than growth had influenced informality in recent years.[4] Second, the informality rates of the groups included in the reform changed after its implementation. The standardized informality rate of workers that earn between one and ten minimum wages decreased significantly after the reform, when compared to the informality rate of the workers outside this bracket, as shown in figure 1.

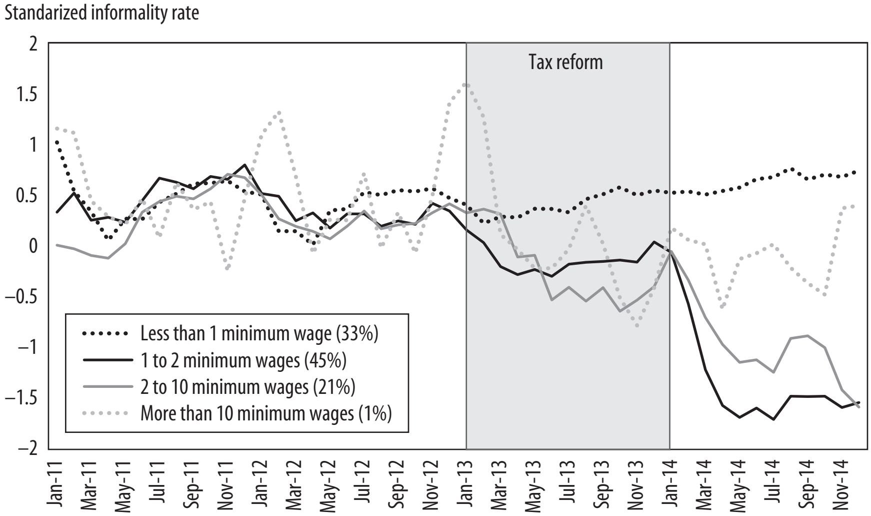
This paper formally estimates how much of the reduction in the informality rate was due to the tax reform, isolating the impact of other macroeconomic and regulatory variables. A natural approach to perform this task is to apply a difference-in-differences (DID) methodology using a household survey panel. The change in the difference in the informality rate of workers affected by the reform and those who were not provides an estimate of the impact of the reform, netting the change in macroeconomic conditions that affected workers included in the reform and those who were not. The difference-in-differences technique has been widely used in the labor market. One of the best-known

2. The new profit tax is the Contribución sobre la Renta para la Equidad y el Empleo (CREE).

3. All the data regarding informality in this section are based on the Integrated Household Survey (*Gran Encuesta Integrada de Hogares,* GEIH) compiled by the National Administrative Department of Statistics (DANE), and they use the legal measurement of informality. Details are provided below, in the data description section.

4. The output gap is from Fedesarrollo; both correlations are significant at 1 percent. This estimation makes use of the firm definition of informality (see the data description section) since the series are longer. The correlation between the firm definition and the legal labor market definition across time is 0.93.

FIGURE 1. Informality Rate by Wage Levels[a]



Source: Authors' calculations, based on the GEIH.

a. Standardized series, where the informality rate is defined using the legal definition. The sample covers the thirteen metropolitan areas. In the legend, the numbers in parentheses are the shares of each group in the working population.

papers is by Card and Krueger, who analyze the impact of the increase in the minimum wage in New Jersey on employment in fast food restaurants.[5] On informality, Bergolo and Cruces also apply a difference-in-differences technique to estimate the impact of an increase in health services coverage for dependent children of private sector salaried workers on informality rates.[6] Slonimczyk, in a very similar setting to the one presented in this paper, finds that a 17 percent reduction in payroll taxes in Russia in 2001 reduced the informality rate by between 2.5 and 4.0 percent.[7]

Since the Colombian Household Survey does not have a panel structure, we estimated the impact of the reform by using a matching difference-in-differences (MDID) method with repeated cross-sections, as suggested by Heckman, Ichimura, and Todd.[8] This method simulates an experiment by matching the treated and untreated populations before and after the reform. The mix of difference-in-differences and matching techniques is not widely

5. Card and Krueger (1994).
6. Bergolo and Cruces (2011).
7. Slonimczyk (2012).
8. Heckman, Ichimura, and Todd (1997).

used in the literature. One notable exception is the evaluation of training programs.[9] Encina also uses this method to analyze the impact of the pension reform on the labor participation outcomes in Chile; Villa, Fernandes, and Bosch apply the MDID approach to estimate the impact of behavioral interventions on the self-employed informality rate; and the World Bank uses a synthetic panel to find that a one-percentage-point decrease (increase) in the labor cost ratio (formal to informal) results in a 2.2 percentage point rise (fall) in the fraction of jobs that are registered.[10]

According to our MDID estimations, the 2012 Colombian tax reform can be associated with a reduction of 4.8 percentage points in the informality rate of workers affected by the reform in the thirteen metropolitan areas. This is equivalent to a reduction in the overall informality rate of around 2.1 percentage points, provided that the treated population was 43 percent of the working population in the thirteen metropolitan areas in 2014. The reduction in informality rates was greater for men than it was for women, for urban than for rural workers, and for those with lower levels of education. Our estimated results compare relatively well with previous literature, but the magnitude of the effects is rather on the low side. Antón estimates that the 2012 tax reform in Colombia increased formal employment by between 3.4 and 3.7 percent and decreased informal employment by between 2.9 and 3.4 percent.[11] Similarly, recent studies sponsored by the Inter-American Development Bank (IDB) find that the reform increased the absolute number of formal jobs by between 200,000 and 800,000 (an increase of the number of formal jobs of between 3.1 and 3.4 percent relative to December 2012) and increased wages from 1.9 to 4.4 percent.[12] Previous work on payroll taxes in Colombia finds similar results. Kugler and Kugler find that a 10 percent increase in payroll taxes leads to an increase in informal employment of between 4 and 5 percent, while Mondragón-Vélez, Peña, and Wills find that a 10 percent increase in payroll contributions was correlated with an increased probability of informality ranging between 5 and 8 percentage points.[13]

Our estimated results are robust to different specifications. The data also indicate that before the reform, the outcome of affected and unaffected workers

9. For example, Blundell and others (2004); and Bergemann, Fitzenberger, and Speckesser (2009).

10. Encina (2013); Villa, Fernandes, and Bosch (2015); World Bank (2009).

11. Antón (2014).

12. Steiner and Forero (2016); Kugler and Kugler (2015); Bernal, Eslava, and Meléndez (2016).

13. Kugler and Kugler (2009); Mondragón-Vélez, Peña, and Wills (2010).

evolved similarly, signaling that the divergence in series might be associated with the reform. However, three issues might affect the interpretation of our results. First, recent formalization policies applied to the self-employed might be biasing our results, as self-employment is an important share of the control group and was not affected by the tax reform. To control for this possible bias, we estimated the MDID over a sample of only salaried workers (in both the control and treatment groups). The results showed even higher effects of the tax reform on informality of salaried workers, though those effects were less robust than estimates for the full sample in terms of some of their statistical characteristics. Second, some workers who earned less than the minimum wage before the tax reform may have marginally increased their income coinciding with the tax reform, which would imply that they shifted from the control group to the treatment group, explaining some of our estimated impact. We argue that this possible effect goes in exactly the same direction as the spirit of the reform, which sought a reduction of informality and increased wages paid to workers through a reduction in the tax wedge, making it unnecessary to isolate the impacts. Third, the decrease in payroll taxes after the 2012 tax reform was accompanied by an increase in the minimum wage, which cannot be easily isolated from the payroll tax. Given that the increase in the minimum wage may have reduced the positive impact of the reduction in the payroll tax on the informality rate, our results may be underestimating their total effect.

In sum, beyond these caveats, the reduction in payroll taxes seems to be responsible for some of the recent reduction in Colombian informality rates. This result is important not only for Colombia but for other countries facing high hiring costs and debating whether payroll taxes should be either waived entirely or exchanged for other taxes to reduce informality. To formally present our estimation of the impact of the reform, the next four sections of this paper explain the methodology, present the data, show the results, and discuss the limitations of the estimation. The final section concludes the article.

The DID and MDID Methods

One of the most adequate methodologies for evaluating the impact of the tax reform on informality, while isolating the impact of growth and other macroeconomic variables over time, is the difference-in-differences (DID) method. This method, applied to the informality framework, involves dividing the population into two groups: one affected by the reform, the *treatment* group, and the other unaffected by the reform, the *control* group. The change in the probability of informality within the control group is then compared with the change observed in the probability of informality within the treatment group. By taking the difference between these changes—or the difference in differences—one isolates structural differences between the groups and factors that affect both groups simultaneously, such as macroeconomic conditions, assuming that the impact of unobservable variables on informality is evenly spread between the two groups. The DID equation, over repeated cross-sections, can be written in two ways. The first is the traditional ordinary least squares (OLS) notation:

$$\text{INF}_{it} = \beta_0 + \beta_1 Y_t + \beta_2 T_{it} + \beta_3 \left(T_{it} Y_{it}\right) + \beta_4 X_{it} + u_{it}, \tag{1}$$

where i refers to the individual; $t = 0$ refers to the value of variables before the reform and $t = 1$ to the value of variables after the reform; INF_{it} is a binary variable that takes the value of one if person i at time t is an informal worker and zero if he or she is a formal worker; Y_{it} (year) is a dummy variable that takes the value of zero in the baseline period and a value of one in the period after the reform; T_{it} (treatment) is a dummy variable that takes the value of one if the individual is from the treatment group and zero if not; and X_{it} refers to the observable characteristics of each individual i at time t, as the probability of an individual's being informal changes according to his or her observable characteristics. The second way to write an equation for changes in informality of a repeated cross-section uses Heckman, Ichimura, and Todd's notation:[14]

$$\text{DID} = \begin{bmatrix} E\left(\text{INF}_{it=1} \middle| D_{it=1} = 1, T_{it=1} = 1, X_{it=1}\right) \\ - E\left(\text{INF}_{it=1} \middle| D_{it=1} = 0, T_{it=1} = 0, X_{it=1}\right) \end{bmatrix} - \begin{bmatrix} E\left(\text{INF}_{it=0} \middle| D_{it=0} = 0, T_{it=0} = 1, X_{it=0}\right) \\ - E\left(\text{INF}_{it=0} \middle| D_{it=1} 0 = 0, T_{it=0} = 0, X_{it=0}\right) \end{bmatrix}, \tag{2}$$

where DID, is the difference-in-differences estimation, $D_{it=0}$ is the treatment indicator in the DID setting, and $E(INF_{it=1} | D_{it}, T_i, X_{it})$ is the average outcome by group. Villa clearly illustrates the equivalence between the two notations.[15]

14. Heckman, Ichimura, and Todd (1997).
15. Villa (2016).

Ideally, the DID framework should be applied over a panel of data. If a panel structure is not available, it can also be applied over repeated cross-sections, but the estimations suffer from multiple limitations since the model assumes common time effects across groups and no changes to the composition of each group, which are difficult assumptions to prove in a repeated cross-section.[16] To reduce these limitations, Heckman, Ichimura, and Todd devised the matching differences-in-differences (MDID) method.[17] As in the DID approach, MDID compares the differences in outcomes between the treatment and control groups over time. However, by creating counterfactuals of the control and treatment groups, the MDID approach is able to isolate other effects that may have affected both groups, such that the difference between the two groups before and after the reform provides information about the impact of the reform.[18]

There are multiple ways to find a counterfactual for the individuals in each of the four groups: the control group and the treatment group before and after the reform. In this paper, we follow Heckman, Ichimura, and Todd in using kernel propensity score matching.[19] This method does not take single individuals, but rather uses averages of individuals weighted by their propensity score of being treated. As suggested by Rosenbaum and Rubin, matching on the propensity score is equivalent to matching on covariables, without losing degrees of freedom in the estimation.[20] The kernel method has the advantage of reducing variance and making use of most of the available information. The kernel propensity score MDID can be explained as follows:

$$\text{DID} = \begin{bmatrix} E\left(\text{INF}_{it=1} \middle| D_{it=1} = 1, T_{it=1} = 1\right) \\ -E\left(\text{INF}_{it=1} \middle| D_{it=1} = 0, T_{it=1} = 0\right) W^{c}_{it=1} \end{bmatrix} - \begin{bmatrix} E\left(\text{INF}_{it=0} \middle| D_{it=0} = 0, T_{it=0} = 1\right) \\ -E\left(\text{INF}_{it=0} \middle| D_{it=1} 0 = 0, T_{it=0} = 0\right) W^{c}_{it=1} \end{bmatrix} W^{t}_{it=1}, \tag{3}$$

16. Blundell and Costa Dias (2009). In fact, the model can control for unobservable individual-specific effects and unobservable macroeconomic effects because they cancel one another out, but not for unobservable temporary individual-specific effects.

17. Heckman, Ichimura, and Todd (1997).

18. One of the advantages of this matching over the standard panel is that we can control the change in observable individual characteristics that might affect their probability of being informal over time, such as getting married, being older, or having more education. They also suffer much less from typical panel data problems such as attrition and nonresponse (Verbeek, 2008).

19. Heckman, Ichimura, and Todd (1997).

20. Rosenbaum and Rubin (1983).

where, W_{it} are the kernel weights that estimate the distance between the propensity score of each observation in the treatment group after the reform and the propensity score of each observation in the three other groups (the control groups before and after the reform and with the treatment group before the reform), giving the highest weight to those with propensity scores closest to the treated individual. The treatment group after the reform is assigned a value of one. The expected values of informality are no longer controlled by observed characteristics, since the weights already include this information. Therefore, the second stage of the MDID only estimates an OLS using weights and without covariables. Depending on the researcher's preferences, the procedure can make use of all the propensity score information available or trim it to an area that has available information for both the treatment and control groups (common support).

In sum, by mixing the methodologies, matching and differences in differences, we can control for differences in the composition of the treatment group before and after the treatment. This is not possible in a single DID approach and thus is more suitable for repeated cross-sections.

Data Description

The data set used in this paper is from the official household survey (*Gran Encuesta Integrada de Hogares,* GEIH,) for 2008–15, provided by the Colombian National Administrative Department of Statistics (DANE). This survey collects information for an average of 20,669 households a month, which makes it representative on a monthly basis at the national level and for the thirteen main metropolitan areas. It includes information about household members' income and labor status. Most of the exercises that follow use the thirteen main metropolitan areas, which is more representative than the national sample and more commonly used by the Colombian authorities.[21] However, we also checked the results for the whole sample. This survey does not interview the same individuals across time.

The implementation of the law involved several milestones. Most of the discussions were held between October and November 2012, the Law was

21. The total aggregate GEIH covers twenty-three cities with their rural areas, gathering information on more than 62,000 individuals per month. Of these, just over 23,000 are in the thirteen metropolitan areas, whereas these areas represent 51 percent of the total working population.

passed in December 2012, the first payroll tax reduction became effective in May 2013, and the reform was fully implemented on 1 January 2014.[22] We defined our period of analysis from 2012 (January to December), before the implementation of the reform ($t = 0$), to 2014 (January to December), after the implementation of the reform ($t = 1$).

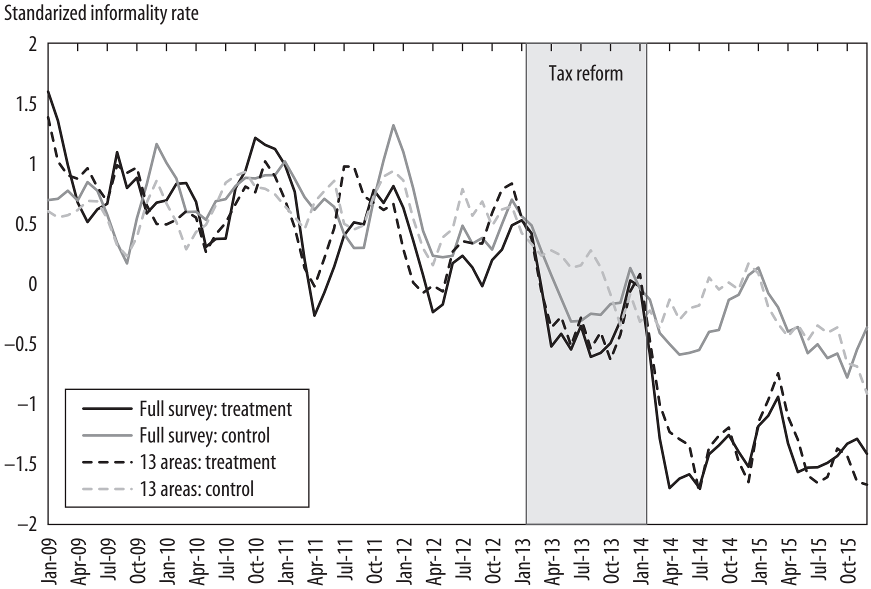
Throughout this analysis, we mostly applied the legal definition of informality, in which informal workers include those who do not make contributions to either health or pension schemes. However, we checked the robustness of the exercises by also applying the so-called firm definition of informality, in which informal workers include those employed in firms with no more than five employees, unpaid family helpers or housekeepers, self-employed workers except for independent professionals and technicians, and business owners of firms with no more than five workers. Results are similar. The results were more conclusive when estimated using the legal definition.

The treatment group in our exercise included all workers that were directly affected by the reduction in payroll taxes. According to the law, this includes workers that earn between one and ten minimum wages, excluding employees of nongovernmental organizations (NGOs) and the government, sole proprietorships, and self-employed workers. After all these exclusions, the reform only covered 43 percent of the working population in the thirteen metropolitan areas in the follow-up period. Our control group includes all other workers. The sample covering the thirteen metropolitan areas provides 219,058 observations in the treatment group and 126,671 observations in the control group. Given that the control group is rather diverse and contains groups that do not share the same logic as the reform's target group—such as the government or the self-employed—we performed another exercise in which we restricted our universe to private salaried workers. Figure 2 presents the standardized series of the treatment and control groups for the thirteen-area sample and the sample that includes rural areas. All cases show a significant reduction of the informality rate of the treatment group over the control group, after the reform was implemented.

According to the MDID setting, all the covariables chosen should affect both the treatment and the outcome variables, without predicting the outcome

22. Of the total reduction of 13.5 percentage points in payroll taxes, 5 percentage points, corresponding to the contributions to the National Adult Training Service (SENA) and the National Family Welfare Institute (ICBF), became effective in May 2013, while the additional reduction of 8.5 percentage points in the health coverage contributions became effective on 1 January 2014.

FIGURE 2. Informality Rate by Wage Levels[a]



Source: Authors' calculations, based on the GEIH.

a. Standardized series, where the informality rate is defined using the legal definition. The sample covers the thirteen metropolitan areas.

perfectly but exogenous to both. Hence, we included the control variables that have the greatest impact on informality in the regression.[23] We did not include the income-related variables or the type of occupation, since they do not satisfy the requirement of being exogenous to the treatment or the anticipation of it. The list of covariables used is the following:

—Gender: We separated women registered as spouses from women registered as heads of household, daughters, and so forth, since the two groups have different preferences for formality;

—Age: We included dummy variables in the regression for workers younger than twenty-five and older than fifty, leaving workers between twenty-five and fifty years old as the base group;

—Education: We included a dummy variable in the regression for workers with primary education or less, another for workers with tertiary education,

23. Fernández and Villar (2016).

and a third for workers who had completed high school, leaving workers with middle and high school studies as the base group;

—City: The equation includes dummy variables for workers who live in the three biggest cities and those who live in border cities where informality often goes hand-in-hand with smuggling;

—Rural/urban: When we used the national sample including rural and urban areas, we included the probability of being in a rural area and in the main thirteen metropolitan areas as a covariable, where the base group is in the less populated urban areas;

—Weights: We include the expansion weight of each individual in the survey as a control variable in the MDID estimation, as it can account for variables that may capture relevant factors—such as where individuals live, their demographic characteristics, and perhaps their availability to respond to surveys—that might interfere with the estimation of informality, but are burdensome to include in the estimations;[24] and

—Months: We included all months, minus one in the estimations to simulate a month-to-month matching.

Tables 1, 2, and 3 present, for 2012, the mean difference between the weighted treated and untreated groups in the thirteen-area sample, the full survey, and the sample of just salaried workers in the thirteen metropolitan areas. The data show significant differences in observable characteristics between the treatment and control groups that make the informality rates impossible to compare directly, justifying the use of a DID approach when panel data are available or, alternatively, an MDID approach when they are not.

Applying the MDID Approach to the Colombian Case

Using the data and the framework explained in previous sections, we performed three main estimations: (1) an MDID estimation for the weighted thirteen-area sample; (2) an MDID estimation for the whole weighted sample; and (3) an MDID estimation for the salaried workers in the weighted thirteen-area sample. In performing these estimations, we applied the kernel (Epanechnikov) propensity score, using weights and no common support (to better approximate the national results and to respect the data generation

24. See Dugoff, Schuler, and Stuart (2014).

TABLE 1. Wald Test on Mean Differences between Treatment and Control Groups: Weighted Thirteen-Area Sample

	Control		*Treatment*		*Adjusted Wald test*		
Variable	*Mean*	*Std. error*	*Mean*	*Std. error*	*Difference*	*(F statistic)*	*Prob. > F*
Informality (legal)	0.76	0.002	0.27	0.003	0.49	22,596	0.000
Women—second earner	0.20	0.002	0.13	0.002	0.07	684	0.000
Women—head or other	0.31	0.002	0.26	0.003	0.05	217	0.000
Less than 25 years old	0.16	0.002	0.17	0.002	−0.01	16	0.000
More than 50 years old	0.26	0.002	0.12	0.002	0.14	2,645	0.000
Elementary education or less	0.26	0.002	0.13	0.002	0.13	2,114	0.000
Tertiary education or more	0.29	0.002	0.41	0.003	−0.11	1,008	0.000
Diploma	0.53	0.002	0.73	0.003	−0.20	3,200	0.000
Big city	0.63	0.002	0.75	0.002	−0.12	2,107	0.000
Border city	0.06	0.001	0.03	0.000	0.03	1,746	0.000
Self-employed	0.63	0.002	0.00	0.000	0.63	81,265	0.000
Salaried	0.18	0.002	0.89	0.002	−0.71	81,111	0.000
Less than 1 min. wage	0.54	0.002	0.01	0.001	0.53	50,934	0.000
More than 10 min. wages	0.05	0.001	0.00	0.000	0.05	2,880	0.000
Between 1 and 10 min. wages	0.41	0.002	0.99	0.001	−0.58	62,270	0.000
No. observations	112,110		61,164				

Source: Authors' calculations, based on the GEIH.

TABLE 2. Wald Test on Mean Differences between Treatment and Control Groups: Weighted Full Sample

	Control		*Treatment*		*Adjusted Wald test*		
Variable	*Mean*	*Std. error*	*Mean*	*Std. error*	*Difference*	*(F statistic)*	*Prob. > F*
Informality (legal)	0.83	0.001	0.33	0.002	0.50	34,055	0.000
Women—second earner	0.19	0.001	0.12	0.002	0.07	1,284	0.000
Women—head or other	0.26	0.001	0.22	0.002	0.04	278	0.000
Less than 25 years old	0.20	0.001	0.17	0.002	0.02	102	0.000
More than 50 years old	0.25	0.001	0.13	0.002	0.12	2,901	0.000
Elementary education or less	0.40	0.002	0.20	0.002	0.20	5,126	0.000
Tertiary education or more	0.20	0.001	0.34	0.002	−0.14	2,857	0.000
Informality (legal)	0.40	0.002	0.64	0.003	−0.24	6,612	0.000
Big city	0.26	0.001	0.48	0.003	−0.22	5,770	0.000
Border city	0.03	0.000	0.02	0.000	0.01	379	0.000
Thirteen metropolitan areas	0.41	0.002	0.65	0.003	−0.24	6,100	0.000
Rural	0.27	0.002	0.13	0.002	0.14	3,105	0.000
Self-employed	0.62	0.002	0.00	0.000	0.62	14,1081	0.000
Salaried	0.14	0.001	0.83	0.002	−0.69	86,242	0.000
Less than 1 min. wage	0.68	0.002	0.02	0.001	0.66	160,624	0.000
More than 10 min. wages	0.03	0.001	0.00	0.000	0.03	3,768	0.000
Between 1 and 10 min. wages	0.29	0.001	0.98	0.000	−0.69	183,785	0.000
No. observations	260,026		100,169				

Source: Authors' calculations, based on the GEIH.

TABLE 3. Wald Test on Mean Differences between Treatment and Control Groups: Weighted Salaried Thirteen-Area Sample

	Control		*Treatment*		*Adjusted Wald test*		
Variable	*Mean*	*Std. error*	*Mean*	*Std. error*	*Difference*	*(F statistic)*	*Prob. > F*
Informality (legal)	0.61	0.005	0.21	0.003	0.40	4,605	0.00
Women—second earner	0.17	0.004	0.13	0.002	0.04	69	0.00
Women—head or other	0.34	0.005	0.27	0.003	0.07	162	0.00
Less than 25 years old	0.32	0.005	0.19	0.002	0.13	567	0.00
More than 50 years old	0.13	0.004	0.09	0.002	0.03	75	0.00
Elementary education or less	0.15	0.004	0.12	0.002	0.03	66	0.00
Tertiary education or more	0.39	0.005	0.41	0.003	–0.02	14	0.00
Diploma	0.65	0.005	0.74	0.003	–0.09	262	0.00
Big city	0.68	0.004	0.75	0.002	–0.07	292	0.00
Border city	0.05	0.001	0.03	0.001	0.02	197	0.00
Less than 1 min. wage	0.72	0.005	0.01	0.001	0.71	18,947	0.00
More than 10 min. wages	0.05	0.003	0.00	0.000	0.05	333	0.00
Between 1 and 10 min. wages	0.23	0.005	0.99	0.001	–0.76	25,251	0.00
No. observations	19,673		53,799				

Source: Authors' calculations, based on the GEIH.

process) and month/city clusters (to reduce the problem of autocorrelation), as suggested by Bertrand, Duflo, and Mullainathan.[25]

Table 4 presents the results of these estimations.[26] In the thirteen metropolitan areas (column 1), the control group had an informality rate of 71.6 percent before the tax reform (2012), which decreased to 71.3 percent after the reform (2014). The treatment group reduced its informality rate from 28.5 percent to 23.5 percent. The difference between the control and the treatment groups was –43.1 percentage points at baseline and –47.8 percentage points at follow-up, meaning that the difference-in-differences estimator is –4.8 percentage points. This indicates that the reform can be associated with a reduction of the

25. Bertrand, Duflo, and Mullainathan (2004). According to the authors, the standard errors in difference-in-differences estimations are underestimated and inconsistent due to severe serial-correlation problems caused by three factors: long time series, serially correlated outcome variables, and treatment variables that change very little within groups. However, we argue that this serial-correlation problem is less severe in our data, because we use only two periods: before and after the payroll reform.

26. The Stata code that we used to apply the MDID was designed by Villa (2016). Table A1 in the appendix presents the estimation of the propensity score for being treated, used for the matching procedure in the MDID and the thirteen-areas aggregate. We also estimated all the exercises using the firm definition of informality, with similar results. These results are available on request.

TABLE 4. MDID Matching Results[a]

Statistic	*(1)*	*(2)*	*(3)*
Mean control *t* (0)	0.716	0.735	0.573
Mean treatment *t* (0)	0.285	0.315	0.220
Mean control *t* (1)	0.713	0.722	0.593
Mean treatment *t* (1)	0.235	0.263	0.189
Difference in differences (p.p.)	**–4.78*****	**–3.97*****	**–5.14*****
Standard error	(0.00595)	(0.00421)	(0.01430)
R squared	0.210	0.195	0.152
Treated population in 2014 (% of total)	43.0	32.4	78.2
Impact on relevant informality rate	–2.1	–1.6	–3.8
No. observations	345,729	716,914	149,709
Control 2012	112,110	260,026	19,673
Control 2014	106,948	249,071	16,655
Treatment 2012	61,164	100,169	53,799
Treatment 2014	65,507	107,648	59,582

*** Statistically significant at the 1 percent level.

Source: Authors' calculations, based on the GEIH.

a. The dependent variable is informality (based on the legal definition). Baseline is 2012; follow-up is 2014. All the estimations use Epanechnikov kernel matching weights, no common support, and month/city clusters.

total informality rate of 2.1 percentage points, considering that the weighted participation of the control group in the population is 43 percent.

Table 4 also shows the exercises over the whole sample (column 2) and over salaried workers only in the thirteen-area sample (column 3). The estimated impact on the treatment group using the full national sample is lower than the thirteen-areas estimate (–4.0 versus –4.8 percentage points), a result that can be explained by a lower impact of the reform on the rural population. The higher impact obtained in the salaried thirteen-area sample versus the whole thirteen-area sample (–5.1 versus 4.8 percentage points) might be related to the positive results of the monitoring and control policies applied to self-employment in recent years. However, care should be taken with these two estimations (columns 2 and 3), since the validity of the assumptions required by the model proved to be weaker in these cases, as discussed in the next section.

We also performed the MDID exercise by gender, education, and low-income status, using the thirteen-area sample specifications to allow comparisons (see table 5). According to the results, all women and men with a tertiary education tend to be less affected by the reform than other population segments. This can be explained by the fact that these two groups tend to show higher levels of informality due to their preferences for flexibility and independence, so their decision to be informal is proportionally less driven by

TABLE 5. MDID Matching Results by Gender, Education, and Income Level[a]

	Gender		*Education*			*Income level*
Statistic	*Males* (1)	*Females* (2)	*Males, primary school or less* (3)	*Males, middle and high school* (4)	*Males, tertiary education or more* (5)	*Males, less than 2 minimum wages* (6)
Difference in differences (p.p.)	−5.01***	−3.10***	−9.78***	−7.51***	−3.16***	−6.80***
Standard error	(0.00751)	(0.00663)	(0.01570)	(0.01080)	(0.00929)	(0.01000)
R squared	0.196	0.214	0.258	0.265	0.123	0.320
No. observations	109,480	102,545	20,126	54,149	35,205	81,136

*** Statistically significant at the 1 percent level.

Source: Authors' calculations, based on the GEIH.

a. The dependent variable is informality (based on the legal definition). Baseline is 2012; follow-up is 2014. All the estimations were applied to workers between twenty-five and fifty years old in the thirteen main metropolitan areas and use Epanechnikov kernel matching weights, no common support, and month/city clusters.

TABLE 6. MDID Matching Results: Robustness to Different Specifications[a]

Statistic	*Original specification (1)*	*No weights (2)*	*No clusters (3)*	*Common support (4)*	*No controls (5)*
Difference in differences (p.p.)	−4.78***	−4.76***	−4.78***	−4.8***	−2.42***
Standard errors	(0.00595)	(0.00589)	(0.00303)	(0.00600)	(0.00600)
R squared	0.210	0.202	0.209	0.210	0.250
No. observations	345,729	345,729	345,729	345,711	345,729

*** Statistically significant at the 1 percent level.

Source: Authors' calculations, based on the GEIH.

a. The dependent variable is informality (based on the legal definition). Baseline is 2012; follow-up is 2014. All the estimations were applied to the thirteen-area sample. The original estimation is specification 2 in table 4, which was calculated using Epanechnikov kernel matching weights, no common support, and month/city clusters. The other specifications vary as indicated in the column headings.

monetary factors.[27] We also reduced the men's sample to workers earning less than two minimum wages. The higher observed impact when reducing the sample (–6.8 versus –5.0 in the full male sample) can be explained by the fact that the reform removed a constraint that was bigger for minimum wage earners than for workers receiving higher levels of income, where wages are more flexible. This result is consistent with the higher impact we found for workers with lower levels of education.

In sum, our results suggest that the reform had a relatively strong impact on the target group. The next section presents the robustness and limitations of these estimations and results.

Robustness of Results

The results obtained in the previous section proved to be relatively robust to the sample used in the estimation. Table 6 shows that these results are also robust to changes in the basic specifications of the MDID, such as the use of weights, clusters, and common support. As expected, for the thirteen-area sample, the weights have an impact on coefficients, the use of clusters affects the standard errors, and common support reduces the number of observations. However, all these changes are minimal. The relatively bigger impact that we found in the equation without controls—which is the same as a DID equation estimated with OLS—corroborates the importance of using MDID instead

27. See Fernández and Villar (2016).

of DID in this specific setting. The remainder of this section analyzes the robustness of these results in terms of the validity of the assumptions behind the MDID model.[28] For this purpose, we refer to the thirteen-area sample, the full sample including rural and small urban areas, and the sample of salaried workers only in the thirteen metropolitan areas.

Parallel Trends

Perhaps the most critical assumption of the MDID approach is parallel trends. This feature ensures that in the post-treatment period, the impact is caused by the reform and not by other factors or trends linked to the fact of belonging to either the treatment group or the control group. According to this assumption, unobservable variables such as growth should affect the outcome variable (informality) of the treatment and control groups in a parallel (but not equal) fashion. In other words, if parallel trends hold, in the absence of the treatment (the tax reform) both populations would have experienced the same time trends, conditional on covariables. Figure 2 (presented earlier) shows that the treatment and control groups behaved similarly before the reform was implemented and diverged after.

A simple OLS regression over the 2009–15 period, simulating a reform for each of these years, allows us to identify the changes in time able to generate significant divergence in the informality rate of the treatment and the control populations.[29] We simulated reforms comparable to the 2012 reforms, which involve three years: pre-reform, implementation, and post-reform. More specifically, we simulated reforms implemented in 2010, 2011, 2012, 2013, and 2014, with a dummy variable one year ahead.[30] Formally, equation 1 can be rewritten as follows:

$$\text{(4)} \qquad \mathrm{INF}_{it} = \beta_0 + \beta_1 Y_t + \beta_2 T_{it} + \beta_3 X_{it} + \sum_{k=2010}^{k=2014} \beta_k \left(T_{it} D_{k+1} \right) + u_{it},$$

where D_{k+1} is a dummy variable that takes the value of one in the period after the reform. Unfortunately, in applying this exercise, we had to control by observable characteristics (or use DID) instead of matching (MDID), because it is not clear how the weights of the matching procedure can be estimated

28. See Blundell and Costa Dias (2009) and Lechner (2011) on MDID assumptions.

29. See Autor (2003).

30. As we are using dummy variables, coefficients should be understood as relative to the missing dummies (2009 and 2010) or relative to reforms implemented in 2008 and 2009.

TABLE 7. Ordinary Least Squares, 2009–15[a]

Explanatory variable	*Thirteen-area sample (1)*	*Full sample (2)*	*Salaried thirteen-area sample (3)*
Constant	4.352***	1.619***	5.756**
	(0.713)	(0.416)	(1.824)
Year	−0.002***	0	−0.002**
	(0.000)	(0.000)	(0.001)
Treatment	−0.425***	−0.394***	−0.348***
	(0.002)	(0.002)	(0.003)
Impact of the reform (p.p.)			
2013–15	−2.608***	−3.979***	−1.083
	(0.346)	(0.295)	(0.568)
2012–14	−2.377***	−4.004***	−1.385**
	(0.337)	(0.293)	(0.495)
2011–13	−0.275	−1.330***	0.201
	(0.327)	(0.291)	(0.429)
2010–12	0.318	−0.761**	0.2
	(0.320)	(0.289)	(0.373)
2009–11	0.165	−0.565*	−0.301
	(0.311)	(0.288)	(0.326)
Summary statistic			
No. observations	1,193,947	2,469,176	509,855
F statistic	21,178	38,271	640
R squared	0.36	0.4	0.26

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.
Source: Authors' calculations, based on the GEIH.
a. The dependent variable is informality (based on the legal definition). Results were controlled by the covariables detailed in the text (data description section). Similar results were obtained when controls were not included. Standard errors are in parentheses.

and included in equation 4. According to the results shown in table 7, in the thirteen-area sample, the 2012–14 reform—which we have considered so far—had a significant impact on the informality rate. The 2013–15 reform was also significant, which indicates that it took some time for the reform to reach full impact after it was implemented. The 2009–11, 2010–12, and 2011–13 reforms were not significant, confirming that the parallel trend assumption holds in the series. This is also the case for the salaried thirteen-area sample. When rural and small cities are included, the coefficients of earlier reforms are significant, but the impact is rather small.

However, the most accurate test to prove parallel trends in an MDID approach is probably the placebo test. For this exercise, the MDID method is applied to any other year with similar external characteristics, faking the

TABLE 8. Placebo Test[a]

Sample and period	*DID (p.p.)*	*Std. error*	*Observations*	*R squared*
Thirteen-area sample				
2012–14	−4.78***	0.006	345,729	0.21
2012–10	0.033	0.005	339,128	0.18
2012–09	−0.27	0.005	333,848	0.18
Full sample				
2012–14	−3.97***	0.004	716,914	0.19
2012–10	−0.86**	0.004	704,921	0.17
2012–09	−1.48***	0.004	692,787	0.17
Salaried thirteen-area sample				
2012–14	−5.14***	0.014	149,709	0.15
2012–10	4.5***	0.014	142,075	0.15
2012–09	1.9*	0.013	139,698	0.14

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.
Source: Authors' calculations, based on the GEIH.
a. Results were controlled by the covariables detailed in the text (data description section). Similar results were obtained when controls were not included.

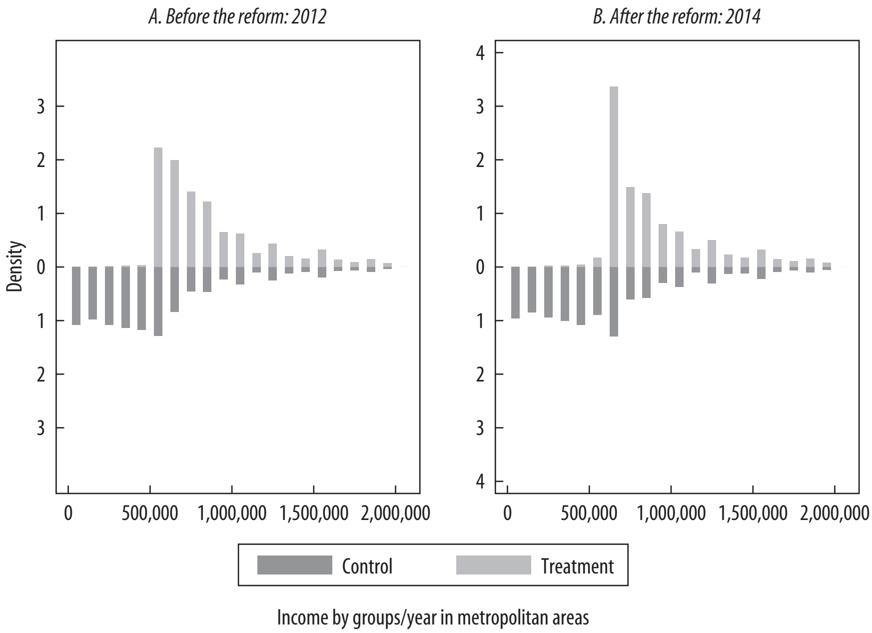
existence of a tax reform or a similar shock, with the expectation that the results will not be affected. We performed this exercise using 2012/2010 and 2012/2009 as alternative periods, simulating a reform that took one year and another that took two years to be fully implemented. In contrast with the years for which we performed our baseline exercise (2014/2012), this alternative should reflect the impact of an nonexistent tax reform. According to table 8, in the thirteen-area sample, we obtained no significant differences between the treatment and control groups in the results on informality. However, results are less clear in the other two samples, confirming the results of the previous exercise.

Exogeneity of the Treatment

A common criticism of the difference-in-differences models with matching, and particularly with MDID with cross sections, is that they have a treated/untreated variable endogenous to the policy implemented. This identification problem, known as Ashenfelter's dip, has been largely analyzed in the literature.[31] It is one of the downsides of using a matching difference-in-differences

31. Abbring and van den Berg (2004); Blundell and Costa Dias (2009); Lechner (2011).

FIGURE 3. Histogram of Wages before and after the Reform: Unweighted Thirteen-Area Sample[a]

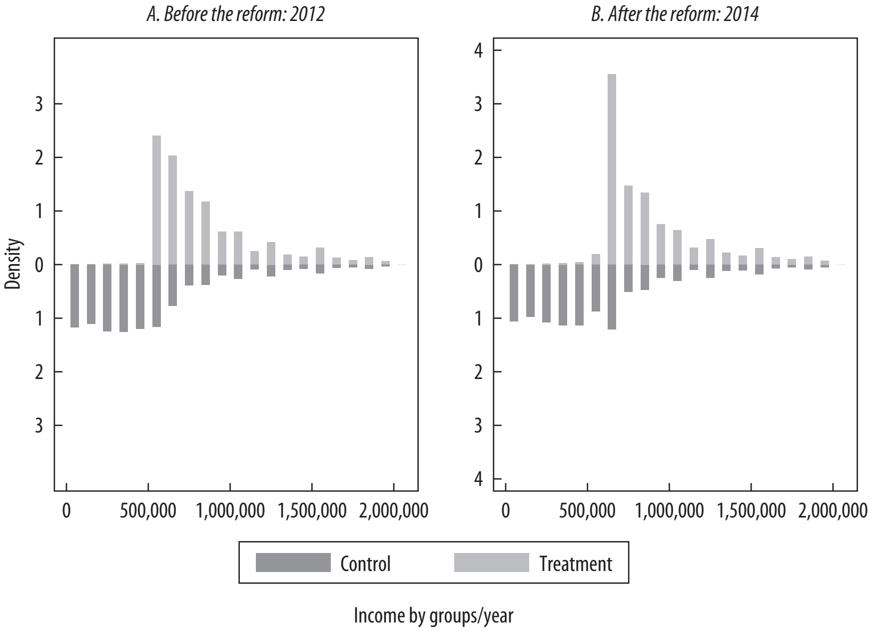


Source: Authors' calculations, based on the GEIH.
a. In 2012, one monthly minimum wage was $566,700 (Colombian pesos); in 2014, $616,000.

method that does not control for unobserved individual-specific shocks that may influence the participation decision. This would be the case, for example, if a benefit program were implemented in two neighboring towns, and individuals migrated to the town where the program was implemented to obtain the benefits.

Two clues indicate that some workers entered the treatment group to obtain the benefits of the reform. The first is that the percentage of formal workers in the treatment group did indeed increase from 41 to 43 percent during the period of analysis. This increase can be explained by a reduction of workers earning less than the minimum wage in the control group. The second is that after the reform, there is an increase in the frequency of workers earning the minimum wage and a decrease in the preceding bracket, as illustrated in histograms of wages (see figures 3, 4, and 5).

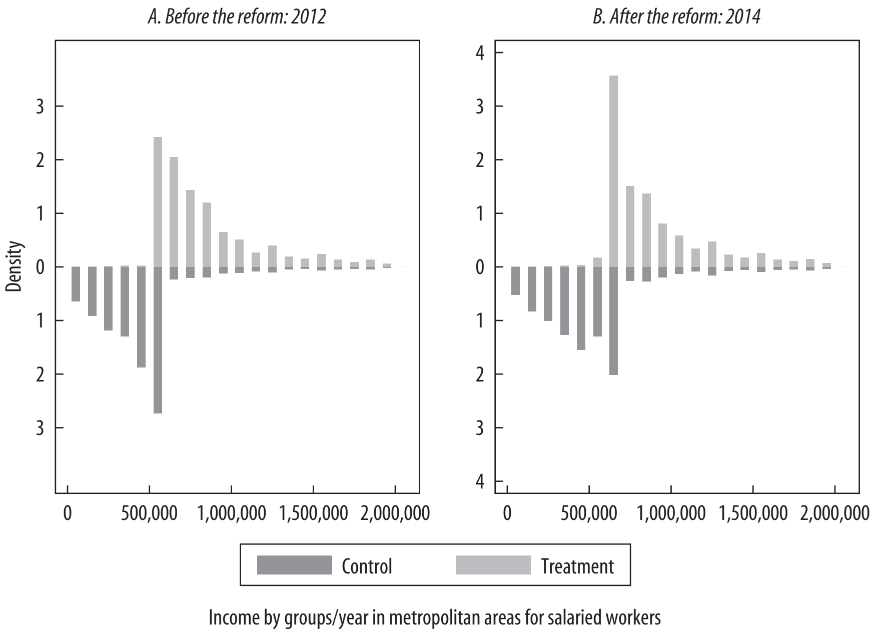
FIGURE 4. Histogram of Wages before and after the Reform: Unweighted Full Sample



Source: Authors' calculations, based on the GEIH.

Unfortunately, we do not have panel data to observe the number of formal workers who transit from control to treatment. However, the direction of the bias that they create goes in the same direction as the spirit of the law. In the case of the lower bound, the impact on the lower threshold was not only positive, but matched the purpose of the reform exactly: to reduce labor costs and thus make it more affordable for firms to pay the minimum wage. It is, in a way, a channel through which the reform reduced informality. This is a desirable result, as quasi-formal workers who worked in the informal sector earning less than the minimum wage moved to the formal bracket and are now likely contributing to the health and pension systems. This problem is very different from cases in which, for example, an individual does not accept a job in order to qualify for unemployment benefits or a higher-wage worker reports an income of less than ten minimum wages in order to obtain access to benefits for low-income families. In the case of Colombia, only 0.8 percent of

FIGURE 5. Histogram of Wages before and after the Reform: Unweighted Salaried Thirteen-Area Sample



Source: Authors' calculations, based on the GEIH.

workers earn more than ten times the minimum wage, so any movements in this segment caused by the reform are not significant.

Quality of the Matching

The robustness of the results also depends on the matching used to create a counterfactual—in other words, on how similar the treatment and control groups are after the matching. The composition of these two groups after the matching (table 9) contrasts with the wide-ranging differences observed before the matching (tables 1, 2, and 3), and shows the effectiveness of the propensity score kernel matching. In most cases, the standardized difference of means is lower than 5 percent, complying with Rosenbaum and Rubin's rule of thumb.[32] However, since we are working with propensity score matching

32. Rosenbaum and Rubin (1985). The only exceptions are the tertiary education variable and the diploma variable in the whole sample, which are on the limit.

TABLE 9. Quality of the Matching[a]

	Thirteen-area sample			Full sample (rural and urban)			Salaried thirteen-area sample		
Explanatory variable	*Mean in treatment*	*Mean in control*	*Standardized difference (%)*	*Mean in treatment*	*Mean in control*	*Standardized difference (%)*	*Mean in treatment*	*Mean in control*	*Standardized difference (%)*
Women (spouse)	0.12	0.12	1.0	0.12	0.1	4.0	0.12	0.12	0.0
Women (other)	0.27	0.27	–0.5	0.26	0.25	3.4	0.29	0.3	–3.9
Under 25 years of age	0.17	0.19	–2.9	0.17	0.16	2.3	0.19	0.2	–2.4
Over 50 years of age	0.12	0.12	0.9	0.13	0.13	–1.3	0.1	0.1	1.3
Primary (–)	0.12	0.12	–1.6	0.14	0.15	–3.5	0.11	0.12	–3.4
Tertiary (+)	0.42	0.39	5.3	0.41	0.38	7.0	0.43	0.41	5.0
Diploma	0.75	0.74	2.3	0.73	0.71	5.3	0.77	0.76	2.4
Big city	0.37	0.36	1.3	0.22	0.21	3.9	0.37	0.36	1.6
Border city	0.08	0.08	0.0	0.07	0.07	0.2	0.08	0.09	–3.9
January	0.08	0.08	0.8	0.08	0.08	1.4	0.08	0.08	2.0
February	0.08	0.08	–0.5	0.08	0.08	–1.0	0.08	0.08	–0.8
March	0.08	0.08	–0.2	0.08	0.08	–0.7	0.08	0.08	–1.1
April	0.08	0.08	–0.4	0.08	0.08	–1.0	0.08	0.08	–0.8
May	0.09	0.09	–0.1	0.08	0.09	–1.0	0.09	0.08	0.3
June	0.08	0.08	–0.9	0.08	0.08	0.4	0.08	0.08	–0.7
July	0.08	0.08	0.5	0.08	0.08	0.2	0.08	0.08	0.5
August	0.08	0.09	–0.5	0.08	0.09	0.0	0.08	0.08	0.9
October	0.08	0.09	–0.3	0.09	0.09	–0.5	0.08	0.09	–1.3
November	0.09	0.09	0.1	0.08	0.08	0.6	0.09	0.08	0.7
December	0.08	0.08	1.2	0.08	0.08	1.2	0.09	0.09	–0.4
Thirteen metropolitan areas	1	1	—	0.61	0.59	3.3	1	1	—
Rural	0	0	—	0.05	0.06	–2.5	0	0	—
Standardized average			0.28			1.0			–0.2

Source: Authors' calculations, based on the GEIH.

a. Dependent variable: Treatment.

instead of one-to-one matching, the average bias of the covariables is what matters the most. In all three samples, the average bias is less than 1 percent, broadly fulfilling Rosenbaum and Rubin's criteria.

Common Support

A key assumption of the MDID procedure is the overlap of the region of common support between the treatment and the control groups. It rules out the perfect predictability of the treatment, given that workers with the same characteristics (X_{it}) might have a positive probability of being both participants and nonparticipants.[33] In other words, we require that $[0 < P(D_{it}) = 1 \mid X_{it}) < 1]$. One way to prove common support is through visual analysis.[34] According to Blundell and Costa Dias, in the MDID model, the propensity score distribution after the reform should be compared with the three control groups (namely, treatment before the reform and control before and after the reform).[35] Figures 6, 7, and 8 show that the propensity score regions of the treated and untreated groups overlap in the three samples, thereby ruling out any concerns about the perfect predictability of the treatment given the observable characteristics.[36]

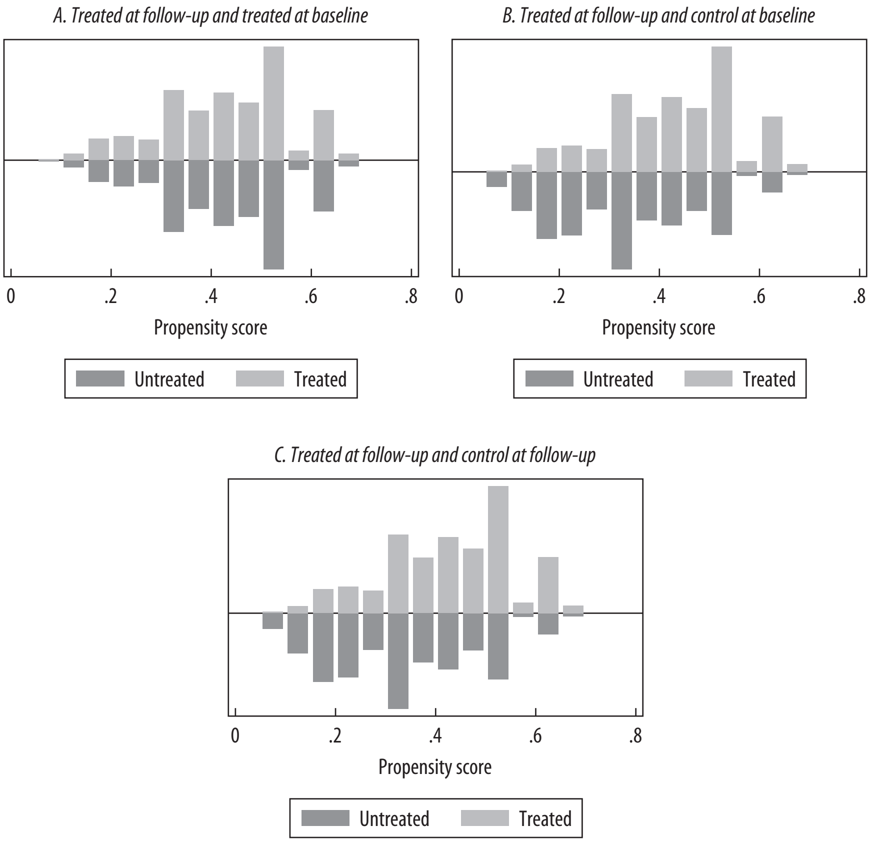
Moreover, when we applied the MDID exercises with and without common support, the number of trimmed observations was minimal (0.01 percent). The differences in the outcome were only reflected in minimal changes in the standard errors, and there were no differences in the coefficients.

Identification of the Unobservable Change

Our discussion in previous sections analyzes the impact of an unobservable change that affected the treatment but not the control group of workers. So far, we have interpreted this change as the reduction in payroll taxes, but it could be related to other regulatory or macroeconomic changes affecting the treatment group but not the control group. One possibility is the general increase in income taxes that accompanied the reduction in payroll taxes. However, this should have affected the treatment and control groups similarly. Moreover, even if the effect on the treatment group was greater, it would not affect our conclusions, given that an increase in income taxes should tend to induce

33. Heckman, LaLonde, and Smith (1999).
34. Caliendo and Kopeinig (2008).
35. Blundell and Costa Dias (2009).
36. The last two distributions are almost equal because the distribution does not change much across years.

FIGURE 6. Propensity Score Distribution: Thirteen-Area Sample

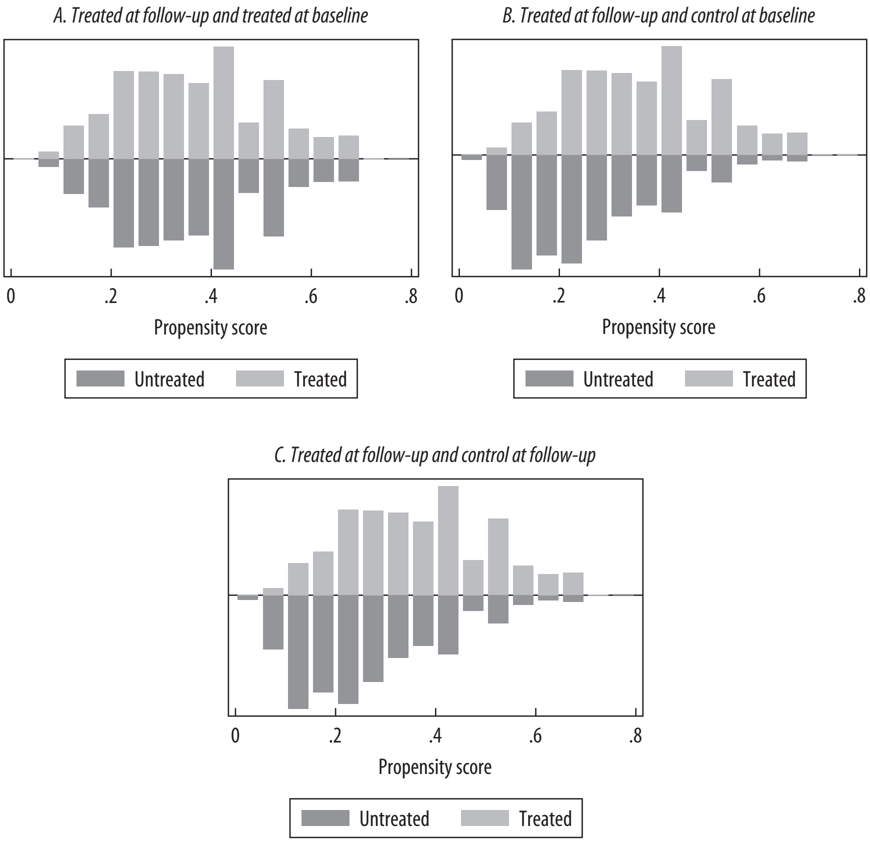


Source: Authors' calculations, based on the GEIH.

informality rather than reduce it. Our results suggest that the reduction in payroll taxes helped to reduce informality even though they were replaced by higher income taxes.[37] Another possibility is the increase in the minimum wage. From 2012 to 2014, the minimum wage increased 1.8 percent annually in real terms (in comparison with 1.1 percent between 2007 and 2011). The

37. Some argue that the creation of the CREE might have offset the impact of the reform, since one tax was replaced by the other. The impact of the reform that we found in the previous section goes against this claim.

FIGURE 7. Propensity Score Distribution: Full Sample



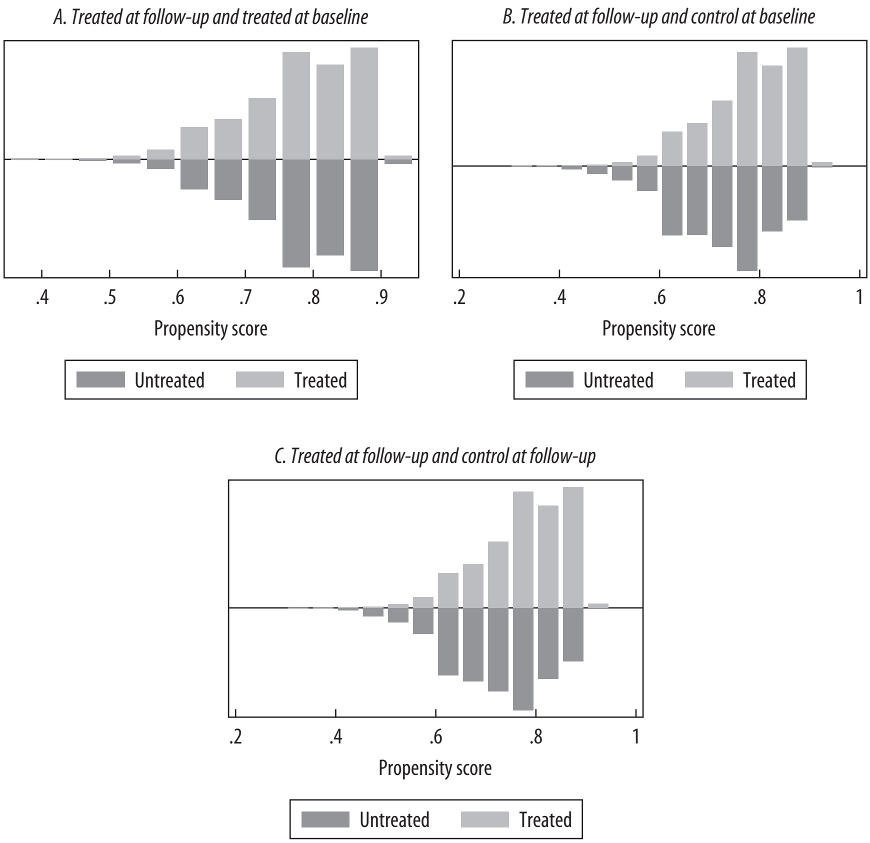
Source: Authors' calculations, based on the GEIH.

impact of the increase is rather difficult to isolate from the reform, as it mostly affected the workers targeted by the reform, but we would expect it to induce an increase in informality. According to Steiner and Forero, the general impact of the tax reform on informality would have been one percentage point greater if the minimum wage had not increased.[38]

In sum, we found that after the 2012 reform, the informality rate of the workers treated by the reform declined more than the informality rate of

38. Steiner and Forero (2016).

FIGURE 8. Propensity Score Distribution: Salaried Thirteen-Area Sample



Source: Authors' calculations, based on the GEIH.

untreated workers. This difference in behavior contrasts with the pattern of previous years, which suggests causality. While the difference could be partially explained by some workers who earned less than the minimum wage and who marginally increased their income to obtain the advantages of the reform, this was the exact purpose of the reform, meaning that this mechanism reinforces the idea that the reform reduced informality. Furthermore, this happened despite the coincidence with elements that should have opposite effects on informality, such as the large increase in the minimum wage. Finally, the results based on the expanded sample covering the whole national population (including smaller cities and rural areas) and the restricted sample of urban salaried workers should be taken with extreme care, as they were less robust, at least with respect to the parallel trend assumption.

Conclusions

This paper attempts to isolate the impact of the 2012 reform on informality from the impact of other macroeconomic and regulatory changes. Given that the Colombian household survey does not have a panel structure, we used a matching difference-in-differences (MDID) methodology to estimate the impact. After the tax reform was implemented in Colombia, the informality rate diminished by 4 percentage points. We argue that approximately half of this reduction is related to the reduction in the payroll contributions adopted by the government. We also found that some of this reduction might be explained by workers who earned less than a minimum wage before the reform and subsequently became fully paid workers to obtain the benefits of the reform. Results were more robust when the sample was restricted to the thirteen main metropolitan areas than when smaller cities and rural areas were included or when the sample was limited to salaried workers, although the impact was higher in this case. The reform also showed a higher impact on men than women and on workers with low levels of education.

Appendix

TABLE A1. Estimation of the Propensity Score at Baseline: Full Sample

Explanatory variable	*Coefficient*	*Std. error*
Constant (B0)	−1.151***	0.033
Women (spouse)	−0.946***	0.019
Women (other)	−0.625***	0.015
Under 25 years of age	−0.397***	0.018
Over 50 years of age	−0.703	0.019
Primary (−)	−0.220***	0.022
Tertiary (+)	0.095***	0.017
Diploma	0.607***	0.020
Big city	0.442***	0.017
Border city	−0.304***	0.018
Thirteen metropolitan areas	0.069**	0.033
Rural	0.013	0.032
Weights	0.046	0.032
January	0.052	0.032
February	0.078**	0.033
March	−0.039	0.032
April	0.059*	0.032
May	0.068**	0.032
June	0.078**	0.032
July	−0.003	0.032
August	0.069**	0.032
October	0.504***	0.017
November	−0.179***	0.024
December	0.000***	0.000
Summary statistic		
No. observations	360,195	
Wald chi squared (23)	14,108.86	
Prob > chi squared	0.0000	
Pseudo *R* squared	0.1044	

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.
Source: Authors' calculations, based on the GEIH.
a. Dependent variable: Treatment.

References

Abbring, Jaap H., and Gerard J. van den Berg. 2004. "Analyzing the Effect of Dynamically Assigned Treatments Using Duration Models, Binary Treatment Models, and Panel Data Models." *Empirical Economics* 29(1): 5–20.

Antón, Arturo. 2014. "The Effect of Payroll Taxes on Employment and Wages under High Labor Informality." *IZA Journal of Labor and Development* 3(1): 1–23.

Autor, David H. 2003. "Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing." *Journal of Labor Economics* 21(1): 1–42.

Bergemann, Annette, Bernd Fitzenberger, and Stefan Speckesser. 2009. "Evaluating the Dynamic Employment Effects of Training Programs in East Germany Using Conditional Difference-in-Differences." *Journal of Applied Econometrics* 24 (5): 797–823.

Bergolo, Marcelo, and Guillermo Cruces. 2011. "Labor Informality and the Incentive Effects of Social Security: Evidence from a Health Reform in Uruguay." Working Paper 62318. Washington: Inter-American Development Bank.

Bernal, Raquel, Marcela Eslava, and Marcela Meléndez. 2015. "Taxing Where You Should: Formal Employment and Corporate Income versus Payroll Taxes in the Colombian 2012 Tax Reform." Bogotá: Universidad de los Andes.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics* 119(1): 249–75.

Blundell, Richard, and Monica Costa Dias. 2009. "Alternative Approaches to Evaluation in Empirical Microeconomics." *Journal of Human Resources* 44(3): 565–640.

Blundell, Richard, and others. 2004. "Evaluating the Employment Impact of a Mandatory Job Search Program." *Journal of the European Economic Association* 2(4): 569–606.

Caliendo, Marco, and Sabine Kopeinig. 2008. "Some Practical Guidance for the Implementation of Propensity Score Matching." *Journal of Economic Surveys* 22(1): 31–72.

Card, David, and Alan B. Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." A*merican Economic Review* 84(4): 772–93.

Dugoff, Eva H., Megan Schuler, and Elizabeth A. Stuart. 2014. "Generalizing Observational Study Results: Applying Propensity Score Methods to Complex Surveys." *Health Services Research* 49(1): 284–303.

Encina, Jenny. 2013. "Reforma de pensiones en Chile: análisis de matching con diferencias en diferencias." *Estudios de Economía* 40(1): 81–95.

Fernández, Cristina, and Leonardo Villar. 2016. "Informality and Inclusive Growth in Latin America: The Case of Colombia." IDS Working Paper 469. Brighton, U.K.: Institute of Development Studies.

Heckman, James J., Hidehiko Ichimura, and Petra E. Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program." *Review of Economic Studies* 64(4): 605–54.

Heckman, James J., Robert J. LaLonde, and Jeffrey A. Smith. 1999. "The Economics and Econometrics of Active Labor Market Programs." In *Handbook of Labor Economics,* vol. 3, edited by Orley C. Ashenfelter and David Card, chap. 31. Amsterdam: Elsevier.

Kugler, Adriana D., and Maurice Kugler. 2009. "Labor Market Effects of Payroll Taxes in Developing Countries: Evidence from Colombia." *Economic Development and Cultural Change* 57(2): 335–58.

———. 2015. "Impactos de la ley 1607 sobre el empleo formal en Colombia." Georgetown University.

Lechner, Michael. 2011. "The Estimation of Causal Effects by Difference-in-Difference Methods." *Foundations and Trends in Econometrics* 4(3): 165–224.

Mondragón-Vélez, Camilo, Ximena Peña, and Daniel Wills. 2010. "Labor Market Rigidities and Informality in Colombia." *Economía* 11(1): 65–101.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70(1): 41–55.

———. 1985. "The Bias due to Incomplete Matching." *Biometrics* 41(1): 103–116.

Slonimczyk, Fabián. 2012. "The Effect of Taxation on Informal Employment: Evidence from the Russian Flat Tax Reform." In *Research in Labor Economics,* vol. 34: *Informal Employment in Emerging and Transition Economies,* edited by Hartmut Lehmann and Konstantinos Tatsiramos, chap. 2. Bingley, U.K.: Emerald Group Publishing.

Steiner Roberto, and David Forero. 2016. "Evaluación del impacto de la reforma tributaria de 2012 a través de equilibrio general." Bogotá: Fedesarrollo.

Todd, Petra E. 1999. "A Practical Guide to Implementing Matching Estimators." University of Pennsylvania.

Verbeek, Marno. 2008. "Pseudo-Panels and Repeated Cross-Sections." In *The Econometrics of Panel Data,* edited by László Mátyás and Patrick Sevestre, pp. 369–83. Berlin: Springer.

Villa, Juan M. 2016. "Diff: Simplifying the Estimation of Difference-in-Differences Treatment Effects." *Stata Journal* 16(1): 52–71.

Villa, Juan M., Danilo Fernandes, and Mariano Bosch. 2015. "Nudging the Self-Employed into Contributing to Social Security: Evidence from a Nationwide Quasi Experiment in Brazil." Working Paper 91877. Washington: Inter-American Development Bank.

World Bank. 2009. "Estimating the Impact of Labor Taxes on Employment and the Balances of the Social Insurance Funds in Turkey." Synthesis Report 44056-TR. Washington.