

Escuela de Economía y Finanzas

# Documentos de trabajo

## Economía y Finanzas

Centro de Investigación  
Económicas y Financieras

No. 13-24  
2013

**Social Assistance and Informality:  
Examining the link in Colombia**

*Ospina.Mónica; Saavedra Caballero, Fabiola*



## **Social Assistance and Informality: Examining the link in Colombia**

**Monica Ospina Londoño**

*Universidad EAFIT*

**Fabiola Saavedra-Caballero**

*Universidad EAFIT*

### **Abstract**

This paper presents evidence on the labor market effects of social assistance programs in the short and medium run. We evaluate the impacts of a Conditional Cash Transfer program (Familias en Acción) on informality in Colombia. We exploit an exogenous shock linked to cash transfer benefits that the government provides to poor people to evaluate the effect of these benefits on informality. We argue that being a beneficiary of social programs may create perverse incentives that drive people towards informality through a substitution effect. Survey data of “Familias en Acción” program was used to identify whether the program had any effect on workers’ labor decisions concerning participation (or the lack thereof) in the informal labor market in Colombia after one and four years after its implementation. We apply matching algorithms and difference-in-differences estimations to evaluate the effect of the program. We find that a worker’s informality condition may be affected by receiving CCT income and by the structure of the colombian health system.

**Key Words:** informality, conditional cash transfers, evaluation

**JEL CODE:** E26, C14, H43

### **Corresponding Author:**

Fabiola Saavedra Caballero

Universidad EAFIT

Carrera 49 N° 7 Sur - 50

Medellín - Colombia – Suramérica

Email: [fsaavedr@eafit.edu.co](mailto:fsaavedr@eafit.edu.co)

Alternative email: [fsaavedra.caballero@gmail.com](mailto:fsaavedra.caballero@gmail.com)

## 1. INTRODUCTION

Different aspects of informality have been studied in the literature, ranging from its definition (Fields, 2011; Levy, 2008; Maloney, 2004; Tokman, 2001; De Soto, 1986, Sethuraman, 1981) to its impact on different areas of the economy, such as reduced public revenues (Turnovsky and Basher, 2009; García-Peñalosa and Turnovsky (2005)), earnings inequality (Rosser, et al., 1999; Ernste and Schneider, 1998) and social policy (Ahmad and Best, 2012; Anton et al., 2011; Levy, 2008; Levy, 2007).

Many studies analyze informality from an exclusionary perspective by considering that it is imposed on workers because of labor market segmentation (Rauch, 1991; De Soto, 1989; Harris and Todaro, 1970). In contrast to this position, we argue that informality might be a worker's choice instead of an imposition ("exit" perspective). Moreover, in line with Levy (2007) and Levy (2008), we believe that informality can also be understood as a result of optimizing decisions that every worker undertakes when he/she has the opportunity to work.

We applied this "exit" perspective to examine Colombia's situation because, over recent decades, the country has reported high levels of informality relative to the rest of Latin America. Studies such as Lopez and Lasso (2012)<sup>1</sup>, Betranou et al. (2009), World Bank (2008) and Uribe and Ortiz (2000)<sup>2</sup> report high informality rates in Colombia since the 1980s. Additionally, Botero (2011) and Gasparini and Tornarolli (2007) find an upward sloping tendency in informality in this country during the 2000s<sup>3</sup>.

Due to this situation, the colombian government has made important efforts to reduce informality. In 1993, a substantial change was implemented through Law 1990 that divided the

---

<sup>1</sup> Lopez & Lasso (2012) used information on the working-age population from household surveys of Colombia's seven capital cities. Their study shows that informal employment (non-salaried with low education level) did not experiment meaningful change between 1984 and 1996 (around 17%), increased due to the Colombian financial crisis at the end of the century (21% highest record), decreased between 2003 and 2007 (18% lowest record) and increased again from 2008 to 2011 (around 22% highest record).

<sup>2</sup> Uribe & Ortiz (2000), using information of the Household National Survey, mention that from the economically active population of the ten most important areas in Colombia, 50.1% were informal workers in 1998, 48.91% were informal workers in 1992, 48.45% were informal workers in 1994, 47.7% were informal workers in 1996, 47.69% were informal workers in 1998, and 48.88% were informal workers in 2000.

<sup>3</sup> Schneider (2012) presents data on Colombia's informality (or shadow economy) measured as a percentage of the official GDP from 1999 to 2010, showing a downward sloping tendency. Nevertheless, the average during that period was 37.7%. It is important to note that informality measurements depend on the definition applied.

health system into subsidized and contributive systems; formal workers would participate in the contributive one, and their employers would pay a percentage of their health and pension contributions<sup>4</sup>. Subsequently, in 2010, Law 1429<sup>5</sup> was enacted for the explicit purpose of reducing firm informality by granting incentives, such as reduced payroll taxes (conditional on hiring vulnerable workers) and reduced formal registration fees<sup>6</sup>, to small enterprises to become formal. However, informality remains high, as 51.2% of the employed population belongs to this labor sector according to DANE (2013)<sup>7</sup>.

In an effort to seek an explanation for the persistence of informality in Colombia from the “exit” perspective, we focused on the benefits that the government provides to the poor for social protection (social protection programs) as a possible source. As theoretical and empirical evidence demonstrates (Urdinola et al., 2009; Gasparini et al., 2006; and Barros, 2006), being a beneficiary of social programs may create perverse incentives. This study highlights those perverse incentives affecting a worker’s labor market sector decision based on Levy (2008).

To achieve our objectives, we employed evaluation data on the “Familias en Acción” program, which allowed us to identify (because of its structure) whether the program had any effect on workers’ labor decisions regarding participation in the informal labor market in Colombia. To accomplish this objective, we applied quasi-experimental estimation techniques (propensity score matching and difference-in-differences estimations) using three different informality definitions. Our results indicate that Familias en Acción program generally increases informality in Colombia.

This first section of the article provides a brief overview of the analysis. A review of the “Familias en Acción” program is presented in section two, while section three assesses different definitions of informality. In section four, we present a literature review that concerns topics related to informality, social protection and conditional cash transfer programs. In section five, we describe the characteristics of our data and develop our main assumptions. Section six

---

<sup>4</sup> The contribution percentages were modified in later years.

<sup>5</sup> Ley 1429 de Formalización y Generación de Empleo

<sup>6</sup> According to Confecamaras (2011), until August 2011, 231.566 enterprises benefited from this law.

<sup>7</sup> The most recent information available is in DANE (2013), based on the GEIH survey - Quarter 4/2012 (October - December) for 13 metropolitan areas in Colombia.

describes the methodologies used to accomplish the impact evaluation and the results of the estimation. Finally, in section seven, we discuss the results and provide some conclusions.

## **2. FAMILIAS EN ACCIÓN**

Familias en Acción (FA) is a social protection program in the form of a Conditional Cash Transfer (CCT). It has been applied by the Colombian government as part of the Red de Protección Social – RPS (Social Support Network) since 2001 with financial support from the World Bank and the Interamerican Development Bank. RPS is a public temporary social safety net that was created in 1999 with the aim of alleviating the effects of an economic recession and fiscal policy adjustments on extremely vulnerable populations in Colombia.

The primary objectives of FA were to safeguard and foster human capital formation among children aged 0 to 17 from poor households through two different subsidies. The first (health and nutrition) was provided to the household on the conditions that children aged below 7 were vaccinated and attended health and nutritional check-ups and that mothers attended informational presentations on health. The second subsidy (education) was provided to children between 7 and 17 who were enrolled in school and attended no less than 80% of classes during the school year.

The eligibility criteria were determined by a proxy-means testing instrument called the System for Identification and Selection of Social Spending Beneficiaries (SISBEN in Spanish), which classifies households into categories according to their characteristics, such as income, education level, and the materials used in home construction, among others. As Castañeda (2005) notes, the application of SISBEN combines geographical targeting and family assessments to determine eligibility for a number of benefits provided by all levels of government. For a household to classify as an FA beneficiary, it must be registered as SISBEN-1, which according to Baez and Camacho (2011a) represents approximately the poorest 20% of the country's population.

The selection of municipalities covered by FA was based on specific characteristics, such as having fewer than 100,000 inhabitants, preferably being located in rural areas, having at least one financial institution, not being the capital of a regional district, having the infrastructure necessary for health and education services, and not being located in the Coffee Region area that

received special assistance as a consequence of the 1995 earthquake (IFS-Econometría S.A., 2004).

To determine the impact of the program, an initial assessment was conducted in 2004, and a second was conducted in 2006. According to official reports (IFS-Econometría S.A., 2004 and DPN, 2008), the general conclusion (of both evaluations) is that the program effectively achieved its objectives, as it increased public expenditures on education, had positive effects regarding growth patterns among rural children and the weight of urban children, and reduced the levels of extreme poverty in rural areas, among others<sup>8</sup>.

### **3. INFORMALITY**

#### **3.1. Defining Informality**

Defining informality is a difficult task; moreover, given the wide variety of concepts used to describe its meaning and evolution, it would be difficult to combine them into a single concept. However, some characteristics agreed on by most definitions are that informality involves economic activities that are not officially measured or registered, the work is generally labor intensive, workers learn skills while working, and entry and exit are easier than in the formal sector. Nevertheless, the most widespread definition of the informal sector is that proposed by the International Labor Organization (ILO) based on the Fifteenth International Conference of Labor Statisticians, which describes it as "...[a group] of units engaged in the production of goods or services with the primary objective of generating employment and incomes to the persons concerned. These units typically operate at a low level of organization, with little or no division between labor and capital as factors of production and on a small scale. Labor relations - where they exist- are based mostly on casual employment, kinship or personal and social relations rather than contractual arrangements with formal guarantees".

The perpetuation and expansion of informality has been attributed to the limited capacity of the formal private sector to generate adequate employment and incomes (Charmes, 1998; Sethuraman, 1997 and Tokman, 1990; quoted in Blunch et al., 2001). However, some studies,

---

<sup>8</sup> There are also studies that analyze the impact of FA on areas other than those for which the program was created, such as intra-household time allocation (Ospina, 2010) and voters' behavior (Nupia, 2011).

such as Maloney (2004) argues, the informal sector should be considered an unregulated sector, analogous to the voluntary entrepreneurial small firm sector found in advanced countries, rather than a residual composed of disadvantaged workers rationed out of good jobs.

In any event, as mentioned in Andrews et al. (2011), “In practice, the definition [of informality] that is most useful depends on the policy concern that motivates the analysis and data availability”<sup>9</sup>. Based on that, we will define informality using a social protection criterion and focusing on a worker’s employment characteristics. This definition will lead us to a basic and intuitive classification that divides workers into salaried or non-salaried workers; nevertheless, the complexity of labor markets in Colombia, as in many countries in Latin America and the world, demands a more disaggregated classification within each group.

Given the remarks above, we followed Levy’s (2008) proposal that goes beyond the salaried and non-salaried framework and explains formality/informality as a result of the intersection of the scope of a regulation’s application (social security) and its enforcement. Additionally, we included Gasparini and Tornarolli’s (2007) perspective that classifies informality according to legality<sup>10</sup> (worker’s labor protection and social security) and productivity (a worker’s skills, firm size and a worker’s income). Finally, from a legalistic perspective, we incorporated some distinctive attributes of the Colombian informal sector, such as enrollment in the Subsidized Health System.

From a social security perspective, the salaried worker can be considered informal if he or she is not enrolled in social security by the hiring firm. While this is illegal, firm owners occasionally negotiate directly with future employees to exchange their social security contributions for a higher salary or “under the table” bonuses. Thus, we are uncertain whether this illegal activity (that implies a lack of social security benefits) is voluntary or compulsory. A reasonable hint to identify this type of informality is the firm’s size; frequently, small firms’ owners declare themselves as self-employed to avoid paying their workers social security. Additionally, it is

---

<sup>9</sup> Contrary to this, Fields (2011) remarks that using terms such as “informality”, “informal work” or “informal economy”, as they are code words that mean different things to different people, impedes communication rather than enhancing it.

<sup>10</sup> As mentioned in Levy (2008), hiring salaried workers but not enrolling them in social security is an illegal act committed by the firm, not by the worker.

very difficult for regulatory institutions to track this irregularity. De Soto (1989) explains this relationship as follows, “higher levels of government rent-seeking or bureaucratic obstacles to legal firm entry will lead to a greater bifurcation of firm sizes; very small informal firms and large formal firms”. Moreover, Charlot et al. (2011) assert that for developing countries, “firms in the informal sector are generally smaller and more numerous compared to those in the formal sector”. For this reason, an “informal worker” will be defined as any person who works for a firm that has fewer than five employees<sup>11</sup>.

From a productivity perspective, the ILO considers informal sector work to be primarily labor intensive and requiring low-level skills; moreover, workers expand their skills on the job. This suggests that a non-salaried and unskilled worker could be characterized as informal. As non-salaried workers do not make social security contributions, regulation becomes irrelevant. Workers with low skills (expressed as a low education level) are assumed to have low incomes and will also have less of an opportunity to compensate for the lack of social security with private services because of income constraints. Therefore, we will assume that non-salaried workers with low skills tend to be informal.

From a legalistic perspective, it must be noted that health services are compulsory in Colombia. Thus, we will use the two participation options a worker has in the General System of Health Services as a defining factor of formality/informality status. According to Law N° 100, passed on December 23<sup>rd</sup>, 1993, all Colombian residents must be affiliated with the General System of Health Services; thus, workers must be enrolled by their employers in the contributive health system, and individuals who lack employment contracts or sufficient economic resources to participate through their own means must be enrolled by the government (subsidized health system). If a worker is enrolled in the Contributive Health System, this implies that, at some point, he/she has been recognized as a “formal” worker by his/her employer. Moreover, we assume there are some payments (payroll discounts) incurred by the worker to receive health services. However, workers enrolled in the Subsidized Health System do not have an employment contract and do not make any payments to receive health services; this is why a

---

<sup>11</sup> We took in DANE’s criteria to specify informal firms’ size.



person who declares to be working but receives subsidized health services from the government can be considered “informal”.

### **3.2. Informality in Colombia**

As mentioned above, informality can be understood in many ways; however, most developing countries follow the ILO’s precepts as a guideline, and Latin America is no exception. Nevertheless, most of the existing literature on this group of countries has used the criterion of a worker having access to social security benefits as the definition of informality (Perry et al., 2007; Gasparini and Tornarolli, 2006; Rofman and Luchetti, 2006 and World Bank, 2008).

Colombia’s statistical office (DANE) follows ILO and PREALC<sup>12</sup> regarding informality criteria and classifies those who are currently employed by a firm with fewer than five employees and those who do not receive any salary for their work (family or other businesses) as informal<sup>13</sup>.

With respect to the evolution of informality in Colombia, over the last two decades, the country belonged to the group with higher informality rates relative to the rest of Latin America, along with Bolivia, Ecuador, Guatemala, Mexico, Nicaragua, Paraguay and Peru<sup>14</sup> (Gasparini and Tornarolli, 2006). These high levels of informality have affected growth in the country; moreover, from a global perspective, informality in Colombia is approximately 20 percentage points above what would be expected, given its GDP per capita (World Bank, 2008).

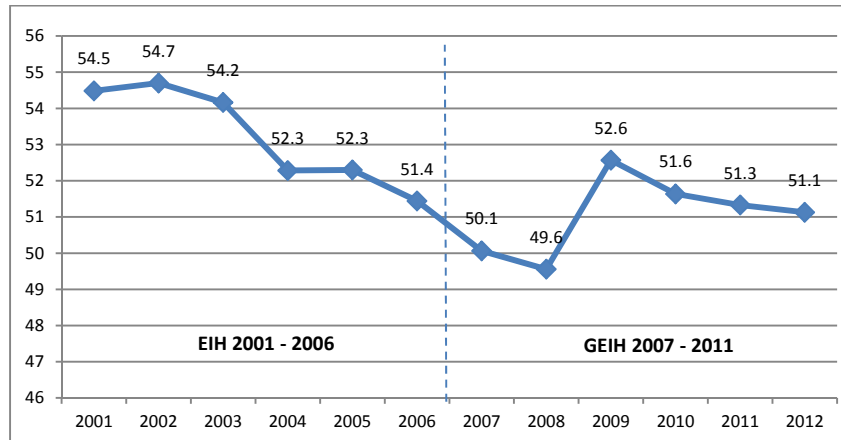
---

<sup>12</sup> Programa Regional de Empleo para América Latina y El Caribe – PREALC.

<sup>13</sup> Guataqui et al. (2011) present a comprehensive study of the components and features of informality in Colombia.

<sup>14</sup> According to Gasparini & Tornarolli (2006), Colombia’s share of informal labor has been close to 60 percent over the past 10 years.

**Figure 1: Informality in Colombia (in percentages)**



**Sources:** Data from DANE, EIH and GEIH

**Notes:** The informality data provided by DANE were extracted from two different surveys. The first contains information from January 2001 until June 2006 called ECH (Continuous Household Survey), and the second contains information from July 2006 until the present, called GEIH (Household Integrated Survey).

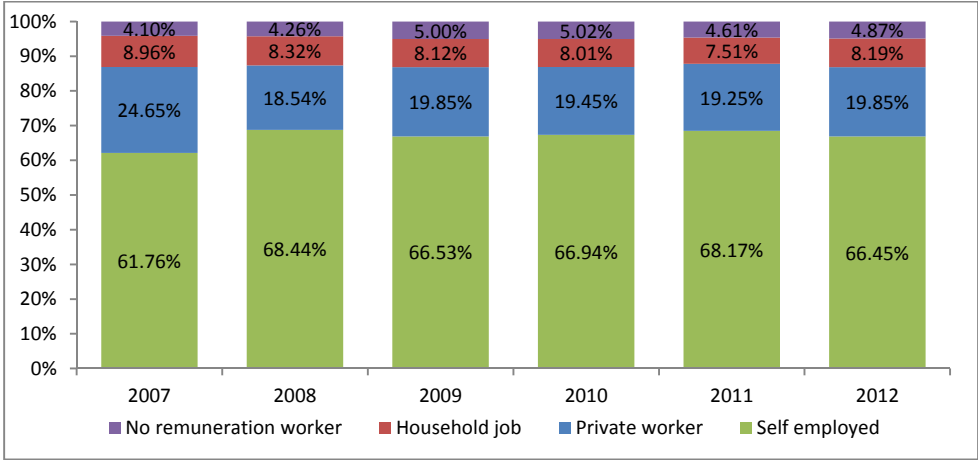
Information for the period 2001 – 2006 is only available from April – June of each year. Therefore, to obtain a single harmonized series, information from 2007 until 2011 also corresponds to the second quarter of each year.

As can be observed in Figure 1, informal workers in Colombia represent an important share of the actively employed population (approximately 50%). Nevertheless, between 2003 and 2008, Colombia's economy was recovering from a recession that, combined with active labor market policies against informality, resulted in a downward sloping trend in informality (with the exception of 2009).

Informality has consistently been a component of the Colombian government's agenda because of its persistence over time and its negative effects. As Perry et al. (2007) note, informality affects workers' welfare, firm productivity and growth, and fiscal sustainability.

With respect to the structure of the Colombian informal labor force (Figure 2), it can be observed that most of these workers declared themselves to be self-employed (over 60%), which according to Bernal (2007) can be defined as those who own their own firms but do not have any paid employees.

**Figure 2 Informal Labor Force by Job Types (in percentages)**



Source: Data from DANE - GEIH

A number of key socioeconomic factors common among informal workers in Colombia are highlighted by the World Bank (2008). These factors include greater participation of men in the informal sector, similar distributions of women and men in the informal labor force by job type, age ranges where there is a greater incidence of informality (from 15 – 18 years and 45 and above) and a higher likelihood that young people will begin their working lives as informal workers.

#### **4. LITERATURE REVIEW**

Informality can be understood from different perspectives, but the “exit” and “exclusion” perspectives seem to be the most appropriate to explain the connection between informality and social protection, as these perspectives consider the different relationships between the state and economic agents. As Perry et al. (2007) note, the “exclusion” of workers from the benefits provided by the government, which can be considered involuntary informality, occurs due to labor market segmentation, burdensome entry regulations for small firms (De Soto, 1989), and excessive tax and regulatory burdens for large firms. Conversely, the “exit” perspective, based on Hirschman (1970), holds that workers, firms and families make their informality decision based on implicit cost-benefit analyses that incorporate the quality of the state’s service provision and its enforcement capacity.

##### **4.1.Social Policy and Informality**

We will focus on the “exit” perspective, which allows<sup>15</sup> workers to consider their current situation when deciding whether to participate in the informal labor market. Some of the reasons that workers might choose informality are related to their perceptions of their own skills<sup>16</sup> and the high mobility that informal workers may have (Hirschman, 1970).

Unskilled workers who partially pay for social protection (formal workers with low skills) directly and implicitly pay for this protection through lower formal wages, and they may find

---

<sup>15</sup> The “exclusion” perspective can be considered an imposition because it does not give economic agents any chance of making decisions regarding their situation. However, the “exit” (meaning “leaving without trying to fix things”, according to Hirschman (1970)), introduces informality as an option that workers choose based on cost-benefit analyses.

<sup>16</sup> This justifies including the perspective advanced by Gasparini and Tornarolli (2007) when defining informality.

that what they expected to obtain in the formal jobs for which they are qualified does not outweigh the forgone current consumption or greater flexibility and earnings they may obtain as informal workers. This is especially true when they have social protection alternatives through access to universal or noncontributory programs or through private means. Larger firms or skilled professionals may decide to underreport their operations and incomes, balancing private gains from tax evasion with the low detection risks resulting from poor enforcement (Perry et al., 2007).

Based on these considerations, we will follow Levy (2008), which is based on the assumption that labor markets in Latin America are not segmented<sup>17</sup>; hence, as in De Soto (1987) and Hirschman (1970), this perspective considers that informality might be the result of workers' decisions. Levy's (2008) main contribution is the inclusion of social protection as a determinant of informality<sup>18</sup> and argues that social policy induces informality for four different reasons: 1) non-salaried workers are excluded from social security; 2) social security functions like a tax on salaried labor; 3) social protection functions like a subsidy to non-salaried labor; and 4) firms and workers in salaried relationships engage in illegal behavior by generating illegal (informal) salaried labor.

These four conclusions are extracted from the theoretical model presented in Levy (2008) and Levy (2007), which permits us to analyze labor market dynamics while including social security and social protection benefits.

We must assume that an agent who works in the formal sector (formal worker) recognizes benefits such as pensions or social security that are prepaid by the hiring firm. Although an informal worker does not obtain those benefits, the government provides him/her other benefits (usually from social protection programs) that are not related to his/her personal characteristics. This is precisely the situation we analyzed, where social protection incentivizes workers to be informal.

---

<sup>17</sup> Perry et al. (2007) concluded that most countries in Latin America (including Colombia) present weak evidence for labor market segmentation.

<sup>18</sup> As mentioned above, Levy's (2008) definition of formality/informality is based on the results on the intersection of the scope of a regulation's application (social security, health or other) and its enforcement.

## **4.2.Previous Evidence**

Several prior studies analyze the influence of social assistance programs on workers' choice of employment sectors around the globe. For Mexico, Levy (2008) suggests that non-targeted social assistance programs may increase informality. Barros (2006), also for Mexico, analyzes the impact of “Seguro Popular” (subsidized health services for informal workers), finding no meaningful evidence of increased informality due to the social assistance program. Galiani and Gertler (2010) find that “70 y más”, a non-contributory Mexican pension program, increases informal employment among elderly people.

For Argentina, Gasparini et al. (2006) find that the “Jefes de Hogar” program (cash transfers to unemployed household heads), disincentives participants from searching for a formal job. Urdinola et al. (2009) evaluate the “Green Card” program (a non-contributory health system for the informal sector) in Turkey and fail to detect any impact of the program (or its rapid expansion) on informality.

Specifically for Colombia, we found a single study (Camacho et al., 2010) that evaluates the Colombian government's expansion of publicly provided health services and finds robust and consistent estimates of an increase in informal employment (between 2 and 5 percentage points) after the expansion.

## **5. DATA**

To apply the theoretical framework developed by Levy (2007) and Levy (2008) and to examine the incentives that social protection generates regarding informality in Colombia, we employ data from the FA program. We selected this database because its structure allows us to confidently apply quasi-experimental techniques; moreover, there is no similar dataset for any social program in Colombia with the magnitude of FA<sup>19</sup>.

In general terms, the FA database includes information from three different surveys: a baseline dataset collected in 2002, an initial follow up completed in 2003, and a second follow up

---

<sup>19</sup> Familias en Acción was the first large-scale evaluation of a social intervention implemented in Colombia (Briceño et al., 2011).

conducted between 2005 and 2006. This information is publicly available and can be disaggregated by household and by individual. To compare the impact of FA over time, we constructed two subsamples; the first matches the baseline survey to the first follow up, and the second combines the baseline survey with the second follow up. These subsamples allow us to explain the evolution of informality given social protection in the short and medium/long run.

Due to the labor market characteristics provided in the survey, we define workers as those who reported spending most of their time working the week prior to the survey or reported having a job (although they were not working at that time). To exclude child labor, only workers aged fifteen years and over were considered (according to Colombian Law N° 515 of 1999). Moreover, due to data collection issues, some municipalities received FA payments before the baseline survey was conducted. This situation caused some data contamination; therefore, only information from municipalities where the baseline survey was applied before FA cash transfers were given was included. Finally, in the second follow up survey, some families that belonged to the control group began receiving FA subsidies. These families were not included in the estimations because their transition from the control to the treatment group caused distortions when characterizing each group during the matching process.

Regarding our definition of informality, as noted in previous sections, we understand informality from a broad perspective and use it in three different ways:

*Informality type 1 (Inf1):* Informal workers are those who are not enrolled in social security by the hiring firm: Assuming that small firms do not pay workers' social security benefits (only in this definition), we can safely argue that people working in firms with fewer than five employees can be considered informal workers.

*Informality type 2 (Inf2):* Informal workers are those workers who are unskilled and do not receive a salary: We follow the productive perspective advanced by Gasparini and Tornarolli (2007) to define informality by including non-salaried workers' skills in our conceptualization.



Due to data constraints, we used non-salaried workers' education levels as a proxy<sup>20</sup>. As the target population of FA are SISBEN 1 families (poor people with low education levels and economic limitations), we considered those non-salaried workers who completed high school and above to be “skilled”.

*Informality type 3 (Inf3)*: Informal workers are those enrolled in the Subsidized Health System: With the aim of accounting for Colombia's particular characteristics in our conceptualization, we decided to use enrollment in the Subsidized Health System as the criterion to define informality. If a worker is enrolled in this system, we assume that he/she prefers to be informal and to continue receiving public health services for free rather than making any contributions to obtain similar private health services as a formal worker. Therefore, if a worker is enrolled in the Subsidized Health System, he/she will be considered informal.

## **6. EMPIRICAL ANALYSIS**

### **6.1. Causal effects: Identification and Estimation**

Based on the theoretical framework of Levy (2007) and Levy (2008), we assessed the impacts that social assistance may have on informality. To this end, we used quasi-experimental estimation techniques to evaluate the FA program's impact on informality in Colombia.

Certain considerations, such as differences in covariate values between informal and non-informal workers, demand special attention during the evaluation procedure. Moreover, the non-random assignment of municipalities to the treatment (FA beneficiaries) or control groups may lead to imbalances across the covariates<sup>21</sup>. To avoid biased estimates of the impacts of informality in naïve comparisons of informal and non-informal workers, we applied propensity score matching to control for this confounding imbalance.

---

<sup>20</sup> We decided to use a proxy because it would be almost impossible for us to measure workers' skills properly (including hard and soft skills) based on the available data.

<sup>21</sup> FA program is targeted at poorest people (SISBEN-1) therefore individuals self-select into the program.

Through propensity score matching, we obtained the average treatment effect on the treated (ATT<sup>22</sup>), which tells us whether workers who are FA beneficiaries would have chosen being informal had they not received FA assistance. Nevertheless, this procedure requires a counterfactual to compare with the obtained results; thus, we estimated it<sup>23</sup> by calculating each worker's propensity to be informal using a representation of his his/her decision through a logistic regression of the binary category treatment/control and then matching these workers with other workers with similar propensities. Every decision is based on a function of the individual's observed characteristics (covariates or confounders) summarized in a propensity score<sup>24</sup>.

As in a randomized experiment, matching techniques balance covariate distributions between treated and non-treated individuals as an identification strategy<sup>25</sup>. Treatment ( $T$ ) is assigned independent of potential outcomes  $Y(i)$ , where  $i = 1$  for informal workers and  $i = 0$  for non-informal workers. Therefore, we expect similar average outcomes if both groups receive the same treatment or if none of them do, which can be represented by the following equations:

$$E[Y(1)|T = 1] = E[Y(1)|T = 0] = E[Y(1)] \quad (1)$$

$$E[Y(0)|T = 1] = E[Y(0)|T = 0] = E[Y(0)] \quad (2)$$

These equations show that the average potential outcome for the treatment group under treatment is equal to the average potential outcome of the control group, had it been treated (equation 1), and that the average potential outcome for the treated group, had it not been treated, is equal to the average potential outcome of the control group with no treatment (equation 2).

---

<sup>22</sup> We decided to use ATT instead of ATE because as Caliendo and Kopeinig (2005) note, in order to estimate ATE, we would have to construct not only the counterfactual of treated had they not been treated (as in ATT), but also a counterfactual of the non treated had they been treated.

<sup>23</sup> According to Rosenbaum and Rubin (1983), it is not possible to know each individual's propensity because that would imply having complete information about selection; therefore, these propensities have to be estimated.

<sup>24</sup> By summarizing each individual's characteristics, propensity score matching solves the "curse of dimensionality" that covariate matching would have because of continuous and/or numerous confounders.

<sup>25</sup> The core framework that the propensity score matching method employs attempts to reproduce that of randomized experiments. Nevertheless, as Lee (2006) mentions, it is important to note two important differences between propensity score matching and randomized experiments. First, propensity score matching only balances the observables between treated and control samples, while randomized experiments balance the distributions of both observables and unobservables. Second, the estimates from propensity score matching can be considered a weighted average of the estimates from many small randomized experiments.

Based on this, the ATT is estimated using the following equation, where  $E[Y(0)|T = 1]$  represents the counterfactual:

$$E[Y(1) - Y(0)|T = 1] = E[Y(01)|T = 1] - E[Y(0)|T = 1] \quad (3)$$

However, the estimation of the ATT would only be correct if treatment were assigned randomly, thus making the outcomes independent. Unfortunately, this was not the case because FA beneficiaries were selected based on specific characteristics; as a consequence, we will use the conditional independence assumption (CIA) that ensures that the distributions of key covariates are balanced across the treatment and control groups<sup>26</sup>.

At this point, we have specified our identification strategy (propensity matching score); however, there are many matching metrics available to achieve our goals. The best matching metric is the one that provides the best balance across our covariates of interest; this is the “nearest neighbor”. This metric considers each treated (control) unit and searches for a control (treated) unit with the closest propensity score. Then, this metric matches each unit with the closest one with respect to the propensity score. We used the variation in this metric that includes replacement, which means that an untreated individual can be used more than once as a match for treated units<sup>27</sup>.

- *Difference-in-differences and Matching*

We combined propensity score matching (PSM) with double-difference or difference-in-differences estimation (DiD) to address the potential for unobserved heterogeneity because DiD method allows the program selection to be based on unobserved variables<sup>28</sup>. As Khandker et al. (2010) note, by jointly applying these techniques, we can address the nonrandom program placement which might bias the program’s effect, besides getting a better match control and project units on preprogram characteristics.

---

<sup>26</sup> We have two main assumptions: 1) The set of characteristics we have chosen includes all relevant differences between the treatment and control groups. 2) The distribution of the characteristics in the control group we selected most resembles the distribution of the treatment group.

<sup>27</sup> By allowing replacement in the nearest neighbor technique, we reduce bias and increase average matching quality. Nevertheless, this also increases the variance of the ATT estimators.

<sup>28</sup> Nevertheless, the DiD method requires that unobserved variables be time-invariant.

This estimator (DiD) compares a before-after estimation of treated individuals with a before-after estimation of non-treated individuals. This means, as mentioned in Bryson et al. (2002), that the DiD estimator can cope with macroeconomic changes or changes in the lifecycle socio-economic status, as long as those changes affect both participants and non-participants similarly.

The equation for the DiD estimator is as follows:

$$Y_{it} = \beta_1 + \beta_2 T_i + \beta_3 P_t + \beta_4 (T * P)_{it} + \varepsilon \quad (4)$$

Where  $Y$  represents the outcome,  $T$  is a dummy that takes the value of 1 if the observation is in the treatment group or 0 if it is in the control group.  $P$  represents a post-treatment dummy that takes the value of 1 if the data are from a follow up or 0 if the data correspond to the baseline.

## 6.2. Estimation and Results

As mentioned above, we performed all estimations using two different data subsets; the first includes information from the baseline and the initial follow up, and the second includes information from the baseline and the second follow up. Therefore, all results are presented for both data subsets in parallel.

As a first step, we made a naïve comparison between the treatment and control groups as a reference for further propensity score matching and DiD estimations. To estimate this naïve difference-in-differences estimator, we used the following equation, which was originally proposed by Ñopo et al. (2002) and modified for our purposes:

$$\begin{aligned} & \textit{Effect FA over infX} \\ & = [((\textit{infX}|T = 1) \textit{after FA}) - ((\textit{infX}|T = 1) \textit{before FA})] \\ & \quad - [((\textit{infX}|T = 0) \textit{after FA}) - ((\textit{infX}|T = 0) \textit{before FA})] \end{aligned} \quad (5)$$

The impact that FA has on each type of informality is the proportion of total beneficiaries (treatment group) of the number obtained as a result of equation 5. These calculations are reported in Table 1.

**Table 1: Naïve DiD estimator:**

**Data subset: Baseline vs. First Follow up**

<b>Informality type</b>	<b>Coefficient</b>	<b>S.E.</b>	<b>T-Stat</b>	<b>pvalue</b>	<b>95% Conf. Interval</b>	
inf1	-0,3582%	0,0061	-0,5900	0,5560	-0,0155	0,0083
inf2	-0,6160%	0,0066	-0,9300	0,3500	-0,0191	0,0068
inf3	-3,7263%	0,0082	-4,5200	0,0000	-0,0534	-0,0211

**Source:** Authors' calculations

S.E. Standard Error

**Data subset: Baseline vs. Second Follow up**

<b>Informality type</b>	<b>Coef</b>	<b>S,E,</b>	<b>T-Stat</b>	<b>pvalue</b>	<b>95% Conf, Interval</b>	
inf1	-0,5711%	0,0064	-0,9000	0,3700	-0,0182	0,0068
inf2	2,6581%	0,0076	3,5100	0,0000	0,0117	0,0414
inf3	0,7881%	0,0092	0,8600	0,3890	-0,0101	0,0258

**Source:** Authors' calculations

S.E. Standard Error

As can be observed in the table, the impact that FA has on informality in data subset BF (short run) is even because all the informality types get reduced. However, when we consider the BS (long run) subset, we observe an uneven behavior, *inf1* decreases in 0.5% while *inf2* and *inf3* increase. Clearly, these naïve comparisons and estimators do not consider the features of the informality types included or the structure of the treatment and control groups (most of the regressions are not statistically significant); therefore, to make this information truly comparable, we applied the propensity score matching technique and difference-in differences estimation in the next step.

The covariates used in the logistic regression and their descriptions are specified in Appendix 1. Although the differences are minimal, each definition of informality required a different equation for each data subset (BF and BS) that was constructed from a set of covariates and then used as input to perform propensity score matching and the difference-in-differences estimations.

In addition to the proper selection of covariates for each equation, there are other latent issues we must address, such as the effectiveness of our matching procedure and the reduction of bias. Therefore, we must focus on how well balanced our covariates are; as D’Agostino and Rubin (2000) note, the propensity scores only serve as devices to balance the observed distribution of covariates across treated and comparison groups. Thus, the success of the propensity score estimation is assessed on the basis of this balance rather than by the fit of the models used to create the estimated propensity score.

Another aspect to consider is the assumption of common support. As noted above, PSM estimation performs under this assumption, which implies that all individuals who lack a match or were poorly matched will be omitted from the ATT estimation. Therefore, we look for a sample with high levels of “on support” observations because the larger the sample from which the impact of FA was estimated, the more relevant and representative the results.

Given these considerations, we performed PSM using the *psmatch2* Stata routine developed by Leuvin and Sianesi (2003). We applied the nearest-neighbor metric with replacement and included Cochrane and Rubin (1973) caliper matching, which allowed us to define a tolerance

level when comparing propensity scores (we set the caliper tolerance at 0.01)<sup>29</sup>. The PSM results obtained are reported in Table 2.

The complementary estimations that include different tolerance levels are presented in Appendix 2; they give evidence of robustness since the results are not seriously affected despite different calipers. Similarly, in Appendix 3 we present the results obtained with different algorithms such as Radius Matching, Kernel Matching and NNMatch with the aim of showing that nearest neighbor algorithm not only presented the best balance, but also, the specification was statistically significant for all cases. Finally in Appendix 4 we present heterogeneity test for each equation<sup>30</sup>.

---

<sup>29</sup> An issue with caliper matching is the difficulty in setting the tolerance level a priori (Smith & Todd, 2005). Therefore, we performed several estimations with different caliper levels; our results (balance and common support) were not particularly sensitive to these variations, and hence, we decided to set the caliper at 0.01, as in most impact evaluation studies.

<sup>30</sup> As can be seen in Appendix 4 there is some evidence of heterogeneity in our estimations.

## Table 2: Propensity Score Matching Results

### Data subset: Baseline vs. First Follow up

Informality type	Sample	Treatment	Control	Difference	S.E.	T-Stat	Support (%)	
inf1	Unmatched	0,1574	0,1600	-0,26%	0,0061	-0,4200		
	ATT	0,1573	0,1373	2,00%	0,0082	2,4500	99,9657	**
inf2	Unmatched	0,1877	0,1949	-0,72%	0,0066	-1,0900		
	ATT	0,1876	0,2061	-1,85%	0,0092	-2,0000	99,9828	**
inf3	Unmatched	0,5445	0,5818	-3,73%	0,0083	-4,4900		
	ATT	0,5445	0,5852	-4,06%	0,0115	-3,5200	99,9657	***

Source: Authors' calculations (Statistically significant: \*10%, \*\* 5%, \*\*\* 1%)

S.E. Standard Error

### Data subset: Baseline vs. Second Follow up

Informality type	Sample	Treatment	Control	Difference	S.E.	T-Stat	Support (%)	
inf1	Unmatched	0,1385	0,1431	-0,45%	0,0064	-0,7000		
	ATT	0,1386	0,1200	1,86%	0,0085	2,1900	99,9508	**
inf2	Unmatched	0,2277	0,2026	2,51%	0,0076	3,3		
	ATT	0,2276	0,2053	2,22%	0,0102	2,19	99,9672	**
inf3	Unmatched	0,5483	0,5402	0,81%	0,0092	0,8800		
	ATT	0,5484	0,5225	2,59%	0,0126	2,0500	99,9179	**

Source: Authors' calculations (Statistically significant: \*10%, \*\* 5%, \*\*\* 1%)

S.E. Standard Error



In the short run (data subset BF), the FA program's impact on informality is negative for informality types 2 and 3. However, in the medium/long run (data subset BS), we obtain preliminary positive results from the ATT estimations for all informality types. This finding could be broadly interpreted as the FA program encouraging informality when workers make their labor sector decisions. Nevertheless, the PSM results cannot be directly interpreted, as we are only using them as a component of the DiD estimation process<sup>31</sup>

Finally, as Heckman et al. (1997) suggest, we performed a DiD estimation that includes PSM specifications and results, as the effectiveness of the DiD estimator depends essentially on the characteristics of unobserved variables that could affect informality. The procedure of DiD estimation is performed only over the "on support" part of the sample after the PSM procedure.

---

<sup>31</sup> The PSM results are a fair approximation of the program's impact, but to correct for potential unobservable differences between the treatment and control groups that may affect the outcome (informality), we conducted a DiD estimation based on the PSM results.

### Table 3: Difference-in-Differences Results

#### Data subset: Baseline vs. First Follow up

<b>Informality type</b>	<b>Impact</b>	<b>S.E.</b>
inf1	-3,8675%	0,0097
inf2	2,5041%	0,0096
inf3	-0,4880%	0,0118

Source: Authors' calculations

#### Data subset: Baseline vs. Second Follow up

<b>Informality type</b>	<b>Impact</b>	<b>S.E.</b>
inf1	-1,2901%	0,0104
inf2	4,9000%	0,0109
inf3	7,3636%	0,0130

Source: Authors' calculations

Table 3 presents estimates of the impact on each type of informality besides the standard errors. In the short run (data subset BF), Type 1 informality decreased by 3,86%, demonstrating that the FA program might disincentive beneficiaries to work in small enterprises. Regarding inf2, there is a 2,50% increase. Finally, inf3 decreased by 0,488%, implying that some beneficiaries in the short run prefer to be part of the contributive health system.

In the medium/long run (represented by the data subset Bs), the FA program has a positive effect on informality types 2 and 3. On the other hand, Type 1 informality decreased by 1,29% due to program participation, which implies a decrease in the number of FA program beneficiaries working in enterprises with fewer than five employees. Regarding non-salaried and unskilled workers, informality (inf2) increased by 4,9% keeping the trend observed in the short run; this can be interpreted as a self-entrepreneurial attitude of workers that has been strengthened by the fact that the CCT acts as a financial safety net that allows them to make more risky decisions such as being informal.

Finally, the enrollment of workers in the Subsidized Health System increased 7,36%. In the short run we observed a small decrease in informality (workers started to contribute), but in the medium/long run workers change their minds and rather being enrolled in the subsidized health system. This might happened because they realize that by keeping their SISBEN status they can have access not only to a subsidized health system, but also to CCTs money. Therefore, by being informal they earn money by working (without making any contribution), have access to health services for free and also receive extra money from FA subsidies.

## **7. CONCLUSIONS**

This paper examines the effect of social protection programs on a broad definition of labor market informality. Social protection programs, such as conditional cash transfers, are known to be a powerful tool to combat inequality and increase welfare in a society. However, CCTs may have significant side effects, especially on health, education and the labor market. Positive income shocks may create incentives for workers to move out of the labor market but may also act as a cushion, driving workers towards risky ventures such as informality.

The social protection program “Familias en Acción” was developed as a social safety net for Colombia’s poorest people after the economic crisis the country suffered during the 1990s. The results this program delivered to its target population (children and teenagers) were positive, such as increasing school enrollment, reducing childhood malnutrition and even discouraging child labor supply (Attanasio et al., 2006).

We examine the causal effect of social assistance programs on worker informality using Familias en Acción (FA). To this end, we computed several matching algorithms and employed nearest neighbor propensity score matching with a caliper and a difference-in-differences estimation for three different definitions of informality. We extend the existing literature on CCT and informality by considering informality from a broader perspective and by using three different definitions of informality. In addition, we exploit several key features of our dataset to examine the short and medium run effects.

Overall, we found that the effect that Familias en Accion had over the number of workers in small firms (inf1) was a contraction. In the short run, the decrease was around 3,86% while in the medium/long run was 1,29 %.

About unskilled workers (inf2), FA fostered informality. In the short run informality increased in 2,50% and almost doubled in the medium/long run (4,90%). We speculate that this reflects workers making risky decisions (starting their own businesses) because they use FA a financial back up for investment. Also FA money could be used as the initial investment to start an informal business.

Finally, about workers enrolled in the Subsidized Health System, in the short run it had a small negative impact (0,488%). Nevertheless, in the medium/long run, the effects of FA fostering informality are higher (7,36%) which could be interpreted as the awareness of workers about the benefits of being informal (no contributions, health services for free and extra money from FA). However, the results for the short and medium/long run effects present heterogeneity.

These findings suggest that social protection programs may had a small impact on the structure of the labor market, but also show at some extent how monetary incentives can distort their labor decisions.

## REFERENCES

- Ahmad, E., Best, M. (2012) “Financing Social Policy in the Presence of Informality” Asia Research Centre Working Paper 54. London School of Economics & Political Science.
- Andrews, D., Caldera Sánchez, A. and Johansson, Å. (2011), “Towards a Better Understanding of the Informal Economy”, OECD Economics Department Working Papers, No. 873, OECD Publishing.
- Anton, A., Hernandez, F. and Levy, S. (2012) “The End of Informality in Mexico? Fiscal Reform for Universal Social Insurance” Mimeo.
- Attanasio, O., Fitzsimons, E., Gomez, A., Lopez, D., Meghir, C. and Mesnard, A. (2006) “Child Education and Work Choices in the Presence of a Conditional Cash Transfer Programme in Rural Colombia” The Institute for Fiscal Studies. Wp06/01.
- Baez, J. and Camacho, A. (2011a) “Assesing the Long-Term Effects of Conditional Cash Transfers on Human Capital: Evidence from Colombia”. Presentation at Grupo de Analisis para el Desarrollo (GRADE). Lima – Perú, May 2011
- Baez, J. and Camacho, A. (2011b) “Assesing the Long-Term Effects of Conditional Cash Transfers on Human Capital: Evidence from Colombia”. Discussion Paper Series No. 5751. The Institute for the Study of Labor (IZA).
- Bernal, R. (2007) “The Informal Labor Market in Colombia: Identification and Characterization”. Economics Faculty. University de los Andes.
- Betranou, F. (coordinator) (2009) “Trabajadores independientes y protección social en America Latina”. International Labour Organization.
- Blunch, N., Canagarajah, S., Raju, D. (2001) “The Informal Sector Revisited: A Synthesis Across Space and Time”. The World Bank
- Botero, J. (2011). “Desempleo e informalidad en Colombia: Un modelo de equilibrio general computable”, Unpublished manuscript. EAFIT University - Colombia.
- Briceño, B., Cuesta, L. and Attanasio, O. (2011) “Behind The Scenes: Managing and Conducting Large Scale Impact Evaluations in Colombia”. Working Paper N° 14 International Initiative for Impact Evaluation 3i.
- Bryson, A., Dorsett, R. and Purdon, S. (2002) “The Use of Propensity Score Matching in the Evaluation of Active Labour Market Policies” Department of Work and Pensions. United Kingdom.

- Camacho, A., Conover, E. and Hoyos, A. (2009) “Effects of Colombia's Social Protection System on Workers' Choice between Formal and Informal Employment”. Economics Faculty. University de los Andes.
- Castañeda, T. (2005). “Targeting Social Spending to the Poor with Proxy-Means Testing: Colombia’s SISBEN System”. Social Protection Discussion Paper No. 0529. The World Bank.
- Charlot, O., Malherbet, F., Terra, C. (2010). “Product Market Regulation, Firm Size, Size, Unemployment and Informality in Developing Economies” Centre Interuniversitaire sur le Risque, les Politiques Économiques et l’Emploi. Working Paper 10-43
- Charmes, J. (1990), “A Critical Review of Concepts, Definitions, and Studies in the Informal Sector”, in: Turnham, D., Salomé, B. and Schwarz, A. (eds.) *The Informal Sector Revisited*. Paris: OECD.
- Cochrane, W. and Rubin, D. (1973) “Controlling bias in observational studies: A Review”, *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, Vol. 35.
- Confecamaras (2011) “Impacto de la formalización empresarial en Colombia” Presentation exposed at the 48th National Assembly of Confecamaras (Confederación Colombiana de Cámaras de Comercio). Colombia
- D’Agostino R.B. Jr. and Rubin, D.B. (2000) “Estimating and using propensity scores with partially missing data”, *Journal of the American Statistical Association*, Vol. 95 749-759.
- De Soto, H. 1989, “The Other Path: The Invisible Revolution in the Third World”. Harper Row, New York.
- Dehejia, R. and Wahba, S. (1999). “Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs”, *Journal of the American Statistical Association*, Vol 94, No. 488, 1053-1062.
- Dehejia, R. and Wahba, S. (2002) “Propensity score-matching methods for nonexperimental causal studies”. *Review of Economics and Statistics* 84 (1), 151-161.
- DPN (2008) “Programa Familias en Acción: Impactos en capital humano y Evaluación beneficio - costo del Programa”. Departamento Nacional de Planeación, Colombia.
- Ernste, D. and Schneider, F. (1998) “Increasing Shadow Economies All Over the World - Fiction or Reality? IZA Discussion paper series, No. 26.
- Fields, G. (2011) “Informality, It’s time to Stop being Alice-in-Wonderland-ish” Unpublished manuscript, Version: February, 2011

Fiszbein, A., Schady N., Ferreira, F., Grosh, M., Keleher, N., Olinto, P. and Skoufias, E. (2009) “Conditional Cash Transfers: Reducing Present and Future Poverty”. World Bank Policy Research Report. The World Bank.

Galiani, S. and Weinschelbaum, F. (2006) “Modeling Informality Formally: Household and Firms”. The World Bank.

García-Peñalosa, C. and Turnovsky, S.J., (2005). Second-best optimal taxation of capital and labor in a developing economy. *Journal of Public Economics* 89, 1045–1074.

Gasparini, L., and Tornarolli, L. (2007) “Labor Informality in Latin America and the Caribbean: Patterns and Trends from Household Survey Microdata.”, Working Paper N°46, CEDLAS.

Guataqui, J.C., Garcia-Sauza, A., Rodriguez-Acosta, M. (2011) “El Perfil de la Informalidad Laboral en Colombia”, Documento de Trabajo No. 95, Facultad de Economía. Universidad del Rosario, Colombia.

Harris, J.R. and Todaro, M. P. (1970) “Migration, Unemployment, and Development: A Two-Sector Analysis” *American Economic Review* 60(1): 126-42.

IFS-Econometría S.A. (2004) “Evaluación del Impacto del Programa Familias en Acción — Subsidio Condicionados de la Red de Apoyo Social” Documentación de la Base de Datos de la Segunda Medición – Primer seguimiento. Unión Temporal IFS - Econometría S.A. – SEI

ILO (2011) “2011 Labour Overview: Latin America and the Caribbean”. ILO Regional Ofce for Latin America and the Caribbean

Khandker, S., Koolwal, G. and Samad, H. (2010) “Handbook of Impact Evaluation” The World Bank

Levy, S. (2006) “Social Policy, Productivity and Growth.” Background paper for the regional study, *Beyond Survival: Protecting Household from Health Shocks in Latin America.*” The World Bank.

Levy, S. (2008) “Good Intentions, Bad Outcomes: Social Policy, Informality and Economic Growth in Mexico” Brookings Institution Press, Washington, D.C.

Leuven, E. and Sianesi, B. (2003). "PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing" Statistical Software Components S432001, Boston College Department of Economics, revised 19 Jul 2012.

Lopez, H. and Lasso, F. (2012). “El Mercado Laboral y el Problema Pensional Colombiano”. Borradores de Economía Núm. 736. Banco de la República – Colombia.

Maloney, W. (2004) “Informality Revisited,” *World Development* 32(7), 1159-78.



Ñopo, H., Robles, M., Saavedra, J. (2002) “Una medición del impacto del Programa de Capacitación Laboral Juvenil PROJoven” Documento de Trabajo N° 36. Grupo de Análisis para el Desarrollo – GRADE

Nupia, O. (2011) “Anti-poverty programs and Presidential election outcomes: Familias en Acción in Colombia”. Series Documentos Centro de Estudios Sobre Desarrollo Económico (CEDE) No. 2011-14. Universidad de los Andes.

Ospina, M. (2010) “The Indirect Effects of Conditional Cash Transfer Programs: An Empirical Analysis of Familias En Acción”. Economics Disertations, Paper No. 67. Department of Economics. Georgia State University.

Perry, G., Maloney, W., Arias, O., Fajnzylber, P., Mason, A. and Saavedra, J. (2007) “Informality: Exit and Exclusion”. The World Bank

Rosenbaum, P. y D. Rubin (1983), “The Central Role of the Propensity Score in Observational Studies for Causal effects”, *Biometrika* 70(1), 41-55

Rosser, B., Rosser, M. and Ahmed, E. (2000), “Income Inequality and the Informal Economy in Transition Economies”. *Journal of Comparative Econometrics* 28, 156 – 171

Schneider, F. (2012) “The Shadow Economy and Work in the Shadow: What Do We (Not) now?” Discussion Paper No. 6423. The Institute for the Study of Labor (IZA).

Sethuraman, S.V., (1981) “The Urban Informal Sector in Developing Countries: Employment, Poverty and Environment (World Employment Programme”. International Labour Organization

Tokman, V. (2001) “Integrating the Informal Sector into the Modernization Process” *SAIS Review*, Volume 21, Number 1, Winter-Spring 2001, pp. 45-60 (Article)

Uribe, J. and Ortiz, H. (2010). “Informalidad laboral en Colombia, 1988-2000: evolución, teorías y modelos”. Colección de Libros de Investigación. Universidad del Valle.

World Bank (2008) “Informality in Colombia: Enhancing Worker Welfare and Firm Productivity” Report No. 42698-CO. Human

## Appendix 1: List of Covariates

Type	Variable name	Description
Continuous	age	Age
	agesq	Age squared
	edumax	Maximum education acquired
	expe	Years of experience
	inc	income
	incsq	income squared
	children	Number of children under 6 years
	youth	Number of people form 7 to 17 years
	elderly	Number of people over 60 in the household
	nmemb	Total number of persons who live in the household
	ch_labor	Number of children working in the household
	youth_labor	Number of young workers in the household
	elderly_labor	Number of elderly people who work
	hh_age	Household head age
	chdeath	Number of children death in the household
	wap	Number of working age people in the household
	fwap	Number of femal working age people in the household
	rooms	Number of rooms in the house
	nmemb	Total number of persons who live in the household
	Dummy	twojobs
married		Marital Status dummy: 1 if the worker has a permanent couple (married or free union)
write		Dummy: 1 if the worker is able to write
head		Head of household dummy: 1 if the worker is head of household or head of household's spouse
hh_married		Dummy: 1 if the head of household is married
hh_falone		Dummy: 1 if the head of household is female and has no couple
hh_fem		Dummy: 1 if the head of household is female
unem		Dummy: 1 if the worker was unemployed the last year
migrate		Dummy: 1 if some household member migrated
read		Dummy: 1 if the worker is able to read
illiterate		Dummy: 1 if the worker is unable to read and write
sav		Dummy: 1 if any household member has savings in a financial institution
debt		Dummy: 1 if any household member has debts
sex		Sex dummy: 1 men, 0 women
own		Dummy: 1 if some household member owns the house
elementary		Dummy: 1 if the worker finished elementary school

high	Dummy: 1 if the worker finished high school
college	Dummy: 1 if the worker finished college
urban	Dummy: 1 urban, 0 rural
extra	Extra Income dummy: 1 if the worker perceived any extra income besides his/her main activity

## Appendix 2: Sensitivity to proximity (calipers)

### Data subset: Baseline vs. First Follow up

Inf1

Caliper	Sample	Treatment	Control	Difference	S.E.	T-Stat	Support	
0.01	Unmatched	0,1574	0,1600	-0,0026	0,0061	-0,42		
	ATT	0,1573	0,1373	0,0200	0,0082	2,45	99,965677	**
0.001	Unmatched	0,1574	0,1600	-0,0026	0,0061	-0,42		
	ATT	0,1577	0,1371	0,0206	0,0082	2,52	99,588124	**
0.0001	Unmatched	0,1574	0,1600	-0,0026	0,0061	-0,42		
	ATT	0,1593	0,1359	0,0233	0,0083	2,82	93,753218	***

Source: Authors' calculations (Statistically significant: \*10%, \*\* 5%, \*\*\* 1%)

Inf2

Caliper	Sample	Treatment	Control	Difference	S.E.	T-Stat	Support	
0.01	Unmatched	0,1877	0,1949	-0,0072	0,0066	-1,09		
	ATT	0,1876	0,2061	-0,0185	0,0092	-2	99,9828	**
0.001	Unmatched	0,1877	0,1949	-0,0072	0,0066	-1,09		
	ATT	0,1879	0,2068	-0,0189	0,0092	-2,05	99,5538	**
0.0001	Unmatched	0,1877	0,1949	-0,0072	0,0066	-1,09		
	ATT	0,1887	0,2100	-0,0213	0,0093	-2,29	94,8344	**

Source: Authors' calculations (Statistically significant: \*10%, \*\* 5%, \*\*\* 1%)

Inf3

Caliper	Sample	Treatment	Control	Difference	S.E.	T-Stat	Support	
0.01	Unmatched	0,5445	0,5818	-0,0373	0,0083	-4,4900		
	ATT	0,5445	0,5852	-0,0406	0,0115	-3,5200	99,9657	***
0.001	Unmatched	0,5445	0,5818	-0,0373	0,0083	-4,4900		
	ATT	0,5448	0,5863	-0,0415	0,0115	-3,6000	99,5366	***
0.0001	Unmatched	0,5445	0,5818	-0,0373	0,0083	-4,4900		
	ATT	0,5506	0,5876	-0,0370	0,0115	-3,2000	93,4443	***

Source: Authors' calculations (Statistically significant: \*10%, \*\* 5%, \*\*\* 1%)

## Data subset: Baseline vs. Second Follow up

Inf1

Caliper	Sample	Treatment	Control	Difference	S.E.	T-Stat	Support
0.01	Unmatched	0,1385	0,1431	-0,0045	0,0064	-0,7000	
	ATT	0,1386	0,1200	0,0186	0,0085	2,1900	**
0.001	Unmatched	0,1385	0,1431	-0,0045	0,0064	-0,7000	
	ATT	0,1387	0,1199	0,0187	0,0085	2,2000	**
0.0001	Unmatched	0,1385	0,1431	-0,0045	0,0064	-0,7000	
	ATT	0,1407	0,1220	0,0188	0,0086	2,1900	**

Source: Authors' calculations (Statistically significant: \*10%, \*\* 5%, \*\*\* 1%)

Inf2

Caliper	Sample	Treatment	Control	Difference	S,E,	T-Stat	Support
0.01	Unmatched	0,2277	0,2026	0,0251	0,0076	3,3000	
	ATT	0,2276	0,2053	0,0222	0,0102	2,1900	**
0.001	Unmatched	0,2277	0,2026	0,0251	0,0076	3,3000	
	ATT	0,2279	0,2054	0,0225	0,0102	2,2100	**
0.0001	Unmatched	0,2277	0,2026	0,0251	0,0076	3,3000	
	ATT	0,2272	0,2068	0,0204	0,0102	1,9900	**

Source: Authors' calculations (Statistically significant: \*10%, \*\* 5%, \*\*\* 1%)

Inf3

Caliper	Sample	Treatment	Control	Difference	S.E.	T-Stat	Support
0.01	Unmatched	0,5483	0,5402	0,0081	0,0092	0,8800	
	ATT	0,5484	0,5225	0,0259	0,0126	2,0500	**
0.001	Unmatched	0,5483	0,5402	0,0081	0,0092	0,8800	
	ATT	0,5482	0,5233	0,0249	0,0126	1,9800	**
0.0001	Unmatched	0,5483	0,5402	0,0081	0,0092	0,8800	
	ATT	0,5496	0,5229	0,0267	0,0127	2,1000	**

Source: Authors' calculations (Statistically significant: \*10%, \*\* 5%, \*\*\* 1%)

## Appendix 3: Matching algorithms

### Data subset: Baseline vs. First Follow up

Inf1

Algorithm	Sample	Treatment	Control	Difference	S.E.	T-Stat	Support
Radius	Unmatched	0,1574	0,1600	-0,0026	0,0061	-0,4200	
	ATT	0,1573	0,1478	0,0094	0,0063	1,4800	99,9657
Kernel	Unmatched	0,1574	0,1600	-0,0026	0,0061	-0,4200	
	ATT	0,1573	0,1481	0,0091	0,0063	1,4500	99,9657

	inf1_first	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]
NNmatch	SATT	0,0045	0,0082	0,5400	0,5890	-0,0117 0,0206

Source: Authors' calculations (Statistically significant: \*10%, \*\* 5%, \*\*\* 1%)

Inf2

Algorithm	Sample	Treatment	Control	Difference	S.E.	T-Stat	Support
Radius	Unmatched	0,1877	0,1949	-0,0072	0,0066	-1,0900	
	ATT	0,1876	0,1898	-0,0022	0,0068	-0,3300	99,9828
Kernel	Unmatched	0,1877	0,1949	-0,0072	0,0066	-1,0900	
	ATT	0,1876	0,1900	-0,0024	0,0067	-0,3600	99,9828

	inf2_first	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]
NNmatch	SATT	-0,0100	0,0088	-1,1300	0,2570	-0,0272 0,0073

Source: Authors' calculations (Statistically significant: \*10%, \*\* 5%, \*\*\* 1%)

Inf3

Algorithm	Sample	Treatment	Control	Difference	S.E.	T-Stat	Support
Radius	Unmatched	0,5445	0,5818	-0,0373	0,0083	-4,4900	
	ATT	0,5445	0,5790	-0,0344	0,0086	-4,0000	99,9657 ***
Kernel	Unmatched	0,5445	0,5818	-0,0373	0,0083	-4,4900	
	ATT	0,5445	0,5809	-0,0364	0,0085	-4,2600	99,9657 ***

	inf3_first	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]
NNmatch	SATT	-0,0517	0,0111	-4,6600	0,0000	-0,0734 -0,0299

Source: Authors' calculations (Statistically significant: \*10%, \*\* 5%, \*\*\* 1%)

## Data subset: Baseline vs. Second Follow up

### Inf1

Algorithm	Sample	Treatment	Control	Difference	S.E.	T-Stat	Support
Radius	Unmatched	0,1385	0,1431	-0,0045	0,0064	-0,7000	
	ATT	0,1386	0,1344	0,0043	0,0067	0,6400	99,9508
Kernel	Unmatched	0,1385	0,1431	-0,0045	0,0064	-0,7000	
	ATT	0,1386	0,1354	0,0032	0,0066	0,4900	99,9672

NNmatch	inf1_second	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]
	SATT	0,0084	0,0085	0,9900	0,3230	-0,0082 0,0250

Source: Authors' calculations (Statistically significant: \*10%, \*\* 5%, \*\*\* 1%)

### Inf2

Algorithm	Sample	Treatment	Control	Difference	S.E.	T-Stat	Support
Radius	Unmatched	0,2277	0,2026	0,0251	0,0076	3,3000	
	ATT	0,2276	0,2014	0,0262	0,0079	3,3300	99,9672 ***
Kernel	Unmatched	0,2277	0,2026	0,0251	0,0076	3,3000	
	ATT	0,2276	0,2015	0,0261	0,0078	3,3500	99,9672 ***

NNmatch	inf2_second	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]
	SATT	.022981	.0100367	2.29	0.022	.0033093 .0426526

Source: Authors' calculations (Statistically significant: \*10%, \*\* 5%, \*\*\* 1%)

### Inf3

Algorithm	Sample	Treatment	Control	Difference	S.E.	T-Stat	Support
Radius	Unmatched	0,5483	0,5402	0,0081	0,0092	0,8800	
	ATT	0,5484	0,5320	0,0164	0,0096	1,7100	99,9179 *
Kernel	Unmatched	0,5483	0,5402	0,0081	0,0092	0,8800	
	ATT	0,5483	0,5331	0,0152	0,0095	1,6000	99,9343

NNmatch	inf3_second	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]
	SATT	.0180565	.012001	1.50	0.132	-.005465 .041578

Source: Authors' calculations (Statistically significant: \*10%, \*\* 5%, \*\*\* 1%)

## Appendix 4: Heterogeneity test (Rosenbaum Bonds)

### Data subset: Baseline vs. First Follow up

infl					inf2					inf3				
Gamma	Q_mh+	Q_mh-	p_mh+	p_mh-	Gamma	Q_mh+	Q_mh-	p_mh+	p_mh-	Gamma	Q_mh+	Q_mh-	p_mh+	p_mh-
1,00	206162,00	206162,00	0,02	0,02	1,00	0,95	0,95	0,17	0,17	1,00	248208,00	248208,00	0,01	0,01
1,05	130337,00	28209,00	0,10	0,00	1,05	179428,00	0,10	0,04	0,46	1,05	353139,00	143349,00	0,00	0,08
1,10	0,58	354568,00	0,28	0,00	1,10	260048,00	0,64	0,00	0,26	1,10	453223,00	0,43	0,00	0,33
1,15	0,05	423934,00	0,48	0,00	1,15	337153,00	141273,00	0,00	0,08	1,15	548918,00	0,47	0,00	0,32
1,20	0,71	490468,00	0,24	0,00	1,20	411057,00	215002,00	0,00	0,02	1,20	640606,00	138956,00	0,00	0,08
1,25	134023,00	554414,00	0,09	0,00	1,25	482037,00	285769,00	0,00	0,00	1,25	728623,00	226698,00	0,00	0,01
1,30	194978,00	615987,00	0,03	0,00	1,30	550331,00	353818,00	0,00	0,00	1,30	813262,00	311023,00	0,00	0,00
1,35	253677,00	675374,00	0,01	0,00	1,35	616154,00	419367,00	0,00	0,00	1,35	894785,00	392194,00	0,00	0,00
1,40	310295,00	732743,00	0,00	0,00	1,40	679694,00	482606,00	0,00	0,00	1,40	973422,00	470448,00	0,00	0,00
1,45	364986,00	788242,00	0,00	0,00	1,45	741117,00	543707,00	0,00	0,00	1,45	104938,00	545995,00	0,00	0,00
1,50	41789,00	842004,00	0,00	0,00	1,50	800574,00	60282,00	0,00	0,00	1,50	112284,00	619023,00	0,00	0,00
1,55	469131,00	894147,00	0,00	0,00	1,55	858199,00	660083,00	0,00	0,00	1,55	119398,00	689701,00	0,00	0,00
1,60	518822,00	94478,00	0,00	0,00	1,60	914114,00	715619,00	0,00	0,00	1,60	126294,00	758182,00	0,00	0,00
1,65	567064,00	993998,00	0,00	0,00	1,65	968428,00	769539,00	0,00	0,00	1,65	132985,00	824604,00	0,00	0,00
1,70	613947,00	104189,00	0,00	0,00	1,70	102124,00	821943,00	0,00	0,00	1,70	139485,00	889093,00	0,00	0,00
1,75	659555,00	108854,00	0,00	0,00	1,75	107264,00	872924,00	0,00	0,00	1,75	145804,00	951763,00	0,00	0,00
1,80	703964,00	113401,00	0,00	0,00	1,80	112271,00	922565,00	0,00	0,00	1,80	151953,00	101272,00	0,00	0,00
1,85	747244,00	117837,00	0,00	0,00	1,85	117153,00	970942,00	0,00	0,00	1,85	157941,00	107206,00	0,00	0,00
1,90	789457,00	122169,00	0,00	0,00	1,90	121917,00	101813,00	0,00	0,00	1,90	163777,00	112986,00	0,00	0,00
1,95	830662,00	126402,00	0,00	0,00	1,95	126568,00	106418,00	0,00	0,00	1,95	169469,00	118622,00	0,00	0,00
2,00	870913,00	130541,00	0,00	0,00	2,00	131113,00	110916,00	0,00	0,00	2,00	175024,00	12412,00	0,00	0,00

Source: Authors' calculations



## Data subset: Baseline vs. Second Follow up

inf1

Gamma	Q_mh+	Q_mh-	p_mh+	p_mh-
1,00	127527,00	127527,00	0,10	0,10
1,05	0,62	193479,00	0,27	0,03
<b>1,10</b>	-0,01	256422,00	0,50	0,01
1,15	0,54	316651,00	0,30	0,00
1,20	111226,00	374411,00	0,13	0,00
1,25	166378,00	429917,00	0,05	0,00
1,30	219409,00	483356,00	0,01	0,00
1,35	270491,00	534893,00	0,00	0,00
1,40	319774,00	584674,00	0,00	0,00
1,45	367394,00	632828,00	0,00	0,00
1,50	413471,00	679472,00	0,00	0,00
1,55	458111,00	72471,00	0,00	0,00
1,60	501413,00	768636,00	0,00	0,00
1,65	543462,00	811333,00	0,00	0,00
1,70	584339,00	852879,00	0,00	0,00
1,75	624115,00	893343,00	0,00	0,00
1,80	662855,00	932789,00	0,00	0,00
1,85	70062,00	971275,00	0,00	0,00
1,90	737463,00	100885,00	0,00	0,00
1,95	773437,00	104558,00	0,00	0,00
2,00	808585,00	108148,00	0,00	0,00

inf2

Gamma	Q_mh+	Q_mh-	p_mh+	p_mh-
<b>1,00</b>	240288,00	240288,00	0,01	0,01
1,05	160544,00	32014,00	0,05	0,00
1,10	0,85	396345,00	0,20	0,00
1,15	0,12	469261,00	0,45	0,00
1,20	0,51	53918,00	0,30	0,00
1,25	118138,00	606357,00	0,12	0,00
1,30	182226,00	671018,00	0,03	0,00
1,35	243928,00	733358,00	0,01	0,00
1,40	303427,00	793555,00	0,00	0,00
1,45	360885,00	851763,00	0,00	0,00
1,50	416448,00	908123,00	0,00	0,00
1,55	470246,00	96276,00	0,00	0,00
1,60	522397,00	101579,00	0,00	0,00
1,65	573006,00	106731,00	0,00	0,00
1,70	622172,00	111741,00	0,00	0,00
1,75	669979,00	116619,00	0,00	0,00
1,80	71651,00	121371,00	0,00	0,00
1,85	761836,00	126005,00	0,00	0,00
1,90	806025,00	130527,00	0,00	0,00
1,95	849138,00	134944,00	0,00	0,00
2,00	891232,00	13926,00	0,00	0,00

inf3

Gamma	Q_mh+	Q_mh-	p_mh+	p_mh-
1,00	14796,00	14796,00	0,07	0,07
<b>1,05</b>	0,52	243885,00	0,30	0,01
1,10	0,34	33537,00	0,37	0,00
1,15	121612,00	42283,00	0,11	0,00
1,20	205267,00	506617,00	0,02	0,00
1,25	28553,00	587038,00	0,00	0,00
1,30	362672,00	664362,00	0,00	0,00
1,35	436936,00	738829,00	0,00	0,00
1,40	508537,00	810651,00	0,00	0,00
1,45	577664,00	880017,00	0,00	0,00
1,50	644491,00	947097,00	0,00	0,00
1,55	709172,00	101204,00	0,00	0,00
1,60	771845,00	107499,00	0,00	0,00
1,65	832636,00	113607,00	0,00	0,00
1,70	89166,00	119539,00	0,00	0,00
1,75	949022,00	125305,00	0,00	0,00
1,80	100482,00	130916,00	0,00	0,00
1,85	105913,00	136379,00	0,00	0,00
1,90	111204,00	141703,00	0,00	0,00
1,95	116363,00	146894,00	0,00	0,00
2,00	121396,00	15196,00	0,00	0,00

Source: Authors' calculations