ACI Research Paper #04-2021

Wage Premium in The Formal Sector: Evidence from IFLS

Xiaochen BIAN

John C. HAM

April 2021

Please cite this article as:

Bian, Xiaochen and John C. Ham, "Wage Premium in The Formal Sector: Evidence from IFLS", Research Paper #04-2021, *Asia Competitiveness Institute Research Paper Series (April 2021)*

Wage Premium in The Formal Sector: Evidence From IFLS

Xiaochen Bian*
Asia Competitiveness Institute,
Lee Kuan Yew School of Public Policy,
National University of Singapore
bian@nus.edu.sg

John C. Ham
New York University Abu Dhabi,
NYU Wagner School of Public Service
jch18@nyu.edu

April 5, 2021

Abstract

This paper studies the wage gap between the formal and the informal sectors in Indonesia using data from the Indonesia Family Life Survey (IFLS). We use propensity score matching and the longitudinal data in IFLS to control for selection into the two sectors. Though the formal sector is believed to offer higher wages on average, the empirical literature has not yet reached an agreement on the existence and size of an earnings premium in the formal sector. We find a significant positive effect of being in the formal sector after controlling for selecting. We find that addressing selection into the formal sector using propensity score matching considerably lowers the premium associated with the formal sector. In addition, we find that this premium varies somewhat on how we use firm size in the definition of the informal and formal sectors. However, the differences are not big, suggesting that one can constructively estimate these differences from the (majority of years in the) IFLS that do not contain information on firm size.

I. Introduction

Though negligible in developed countries, the informal sector accounts for a significant share of the GDP of developing countries. According to La Porta and Shleifer (2008), the informal sector accounts for 30% of the GDP and 46.4% of the total employment in the bottom income quartile of countries, in contrast to 17% of GDP and 10% of total employment in the top income quartile of countries.

The structuralist school in development economics hinges on transforming the developing countries from an agriculture-prevailing economy to a modernized and urbanized economy dominated by manufacturing and service industries. Besides upgrading the dominant industry type of the economy from agricultural to manufacturing and service, another aim of such modernization and urbanization process is to increase the ratio of the labour force whose rights are protected by the law. That is, the goal of the structural reforms of the developing countries' governments usually is to encourage firms to hire workers under contracts that protect the welfare of the employees.

The perceived wisdom of the advocates for the formalization of small business is that jobs in the informal sector have lower wages, worse working conditions, little human capital development, and weak job security compared to those in the formal sector as suggested by Nelson and Bruijn

*A thank you or further information

(2005). In effect, the dualism literature following the Lewis model (Lewis, 1954; Harris and Todaro, 1970; De Soto, 1989) describes the growth of a developing economy in terms of a labour transition between two sectors, the labour-intensive sector and the capital-intensive sector. Therefore, for economic progress to continue in developing countries, it is essential to increase the share of employment in the formal sector, i.e. to maximize the movements from bad jobs to good jobs, or from the informal to the formal sector, while minimizing movements from good jobs to bad jobs, or from the formal to the informal sector.

However, considering that people may be self-selected into different sectors, it is unclear whether such superiority of the formal sector will persist, once controlling for individual characteristics. Hence, whether there is a wage premium in the formal sector compared to the informal sector is still open for discussion in the literature.

In our study, we first explore whether there is selection between the two sectors. In other words, what are the characteristics that make certain individuals prone to work in one sector rather than the other? Second, we adopt the propensity score matching method to check whether selection alone can explain the observed gap between the average earnings of workers in the two sectors; that is, we compare the earnings of individuals working in the formal sector to their counterparts in the informal sector. We also run robustness tests to make sure that our results are not sensitive to the classification of the formal and informal sector.

We organize our paper as follows. Section II reviews the literature on the formal and informal sector in developing countries and the literature on matching. In Section III, we discuss the data we used, namely the Indonesia Family Life Survey (IFLS), by giving a description of the collection process, an account of the data cleaning procedures we used to get the rich set of covariates and outcome variable, and the summary statistics. In Section IV, we describe our econometric approach and our empirical results are presented in Section V.

II. Literature Review

i. Literature Review: Wage Gaps Between the Formal and Informal Sector

Even though researchers have not yet reached an agreement on the definition of the informal sector, it is widely accepted that the informal sector still employs most of work force in developing countries. In Indonesia, according to BPS (the Central Bureau of Statistics), in 2015 the percentage of employment in the informal sectors varies from 51.85% to 57.9% depending on different classifications of the informal sector.[1] Despite the conspicuous share of informal jobs in the total employment in developing countries, the informal sector has not yet received as much attention as it deserves. One of the most important unsettled arguments is whether there exists a wage gap between the formal sector and the informal sector unexplained by individual characteristics.

Papers following the dualistic view of Lewis (1954) describe the informal sector as an inferior sector with risky work conditions, high turnovers, and less remuneration. People in the informal sector are workers who are forced out the formal sector. In the dual sector model, informal workers are queuing for better-rewarded formal jobs, implying that formal jobs are better jobs in terms of earnings.

The opposite strand of literature,such as Maloney (2004), thinks of the informal sector as the voluntary entrepreneurial small firm sector found in developed countries. Staneva and Arabsheibani (2014) even find a significant informal employment wage premium in Tajikistan across the whole earning distribution using a quantile regression decomposition technique.

[1] Our benchmark definition of the informal sector agrees with their estimates. We find that in 1997 and 2000, the informal sector takes up 54.82% and 53.23%.

The consensus between the two strands of literature is that the informal sector is less regulated by the government than the formal sector. From the perspective of workers' benefits, the popular view is that the informal sector is uncovered by minimum wage legislation. There are many studies that focus on the minimum wage effects on the formal and the informal sector. The effect of minimum wage on the formal sector is quite unilateral. Like most of the studies, Maloney (2004) report a positive wage effect of minimum wages for the formal sector in Colombia and a negative employment effect for the formal sector workers. Meanwhile, the employment effects of minimum wage on the informal sector is mixed: on one hand, minimum wage above the original equilibrium in the formal sector might displace formal workers into the informal sector; on the other hand, higher minimum wages increase the incentives of informal workers to quit their informal jobs to look for formal job opportunities, leading to an increase in the unemployment rate and decrease in the informal sector. Moreover, all those studies emphasize people's decision making process in response to an increase in the minimum wage. Some of those studies even provide insights on how people's reactions to the change of minimum wage vary with their social economic status. For instance, Alaniz et al. (2011) show that only formal workers in private firms who earned minimum wage will be laid-off due to the increase of minimum wage in Nicaragua. In contrast, Hohberg and Lay (2015) find that in Indonesia the increase of minimum wage significantly reduces employment in the informal sector but increases overall employment and employment in the formal sector. Despite the consensus among those studies that, workers may switch jobs in between the formal and informal sector to maximize their utilities based on their own characteristics, it is still ambiguous whether wage gap between the two sector remains significantly positive after controlling for workers' demographics and employers' characteristics. For instance, Gong and Van Soest (2002) find that it is the sectoral wage differences that drive male workers' choices between formal employment and informal employment. Many existing studies in this strands conclude that the formal sector wage premium is merely a firm-size wage premium. The formal wage premium decreases drastically, from significantly positive to marginally significant after controlling for firm size in the regression (see, for example, Pratap and Quintin (2006); El Badaoui et al. (2010)).

ii. Literature Review: Matching

Matching methods are becoming more and more prevalent as a means of estimating causal treatment effects for policy evaluation programs, especially in the context of policy evaluation programs with labour market outcomes, the matching method has gained popularity. For instance, Heckman et al. (1997), Heckman et al. (1998), Dehejia and Wahba (1999, 2002), and Smith and Todd (2005) use matching methods to evaluate job training programs in the United States. Similarly, Ham et al. (2011) use propensity score matching to measure the impact of migration within the U.S. on wage growth.

The ideal way to measure the impact of those programs is through the Randomized Controlled Trial (RCT), where all eligible candidates are randomly assigned to the treatment group or the comparison group. The treatment group participates in the program while the comparison group is denied access to the program. Because the treatment is randomly assigned, the difference between the outcomes of the treatment group and the comparison group measures the causal effect of the program net of any confounding factors.

However, in real life, even when random assignment of the treatment is feasible and ethical, the program facilitators may choose to give those who need the program the most the priority of getting treated. Moreover, in many situations, randomization of the treatment is not only unethical but also infeasible. For example, in our study, it is impossible for the government to enforce that all employers provide formal employment contracts considering that the informal sector by its natural evades government regulation and supervision.

In scenarios where a RCT is not realistic, economists are left to work with non-experimental data. The challenge of working with these non-experimental data is to isolate the treatment effect from the confounding factors that cause the selection into the treatment group and affect the outcomes at the same time. Economists have subsequently developed a wide range of estimation techniques to tackle the selection bias. One of the estimation strategies is the instrumental variable approach and the other is the matching method. However, since we can not find any variable which affects earnings only through the employment sector, the instrumental variable approach cannot be used here. Therefore, we can only use matching method to find the earnings premium net of the selection bias.

The literature following Heckman et al. (1997, 1998) has scrutinized how well the matching method performs when compared to the RCT and under what conditions the matching estimates can duplicate the RCT estimates. By comparing the matching estimates to the experimental results, some studies in this strand of literature find that matching estimates mimic the experimental results (Friedlander and Robins, 1995; Heckman et al., 1998; Dehejia and Wahba, 1999; Michalopoulos et al., 2004). Meanwhile, Smith and Todd (2005), who revisit the work by Dehejia and Wahba (2002), find that propensity score matching replicates the experimental benchmark for certain subsamples of the National Supported Work data, but not for other subsamples. This casts doubt on the generalization of the matching method, but at the same time reinforces earlier studies, like Heckman et al. (1997) and Heckman et al. (1998), which emphasize that the data must satisfy certain criteria for matching estimators to perform well. The two crucial assumptions for matching estimates to replicate random assignment is the conditional independence assumption and common support condition. While we can show whether the common support condition is satisfied in our sample, the conditional independence assumption is not testable. The best we can do is to show that the treated and untreated group after matching do not vary statistically significantly in perspectives of all the observable characteristics.

According to this literature (Rosenbaum and Rubin, 1983; Heckman et al., 1997, 1998), only a small proportion of the selection bias was due to unobserved differences between individuals, and propensity score matching controls for observed differences between treatment and comparison group better than other regression analysis. Most importantly, the source of selection bias usually is from “comparing to the wrong people" and “comparing the right people in the wrong proportion" rather than in a few cases where the source of selection bias is actually the differences in the unobserved characteristics.

These studies find that improving data quality can reduce the gap between the experimental estimates and the matching estimates. To be more specific, they find that the matching estimators can successfully replicate experimental estimates only when the data satisfies the following criteria: first, both treatment and comparison groups are from the same data sources, so that the same questionnaire is administered to both groups and therefore their outcomes and control variables are measured in the same way; second, the treatment group and the comparison group live in a common economic environment; last, the data must contain a rich set of variables that affect both program participation and outcome variables. Fortunately, IFLS satisfies all of the three criteria above. We now check the three criteria one by one.

First, regardless of the employment sectors, since our sample consists of only males above 15 years old, all of the respondents answered the same questionnaires within the same year despite slight variations in the questionnaires across waves.[2] Second, regardless of the fact that

[2]Most of the variations of questionnaires across waves are minor. For example, for the job status, the respondents are asked to choose from 6 categories in wave 1 and 2, while in wave 3 and 4 there are two more alternatives for them too choose. However, this variation is only relevant to survey time and is irrelevant to the employment sector.

workers in the formal sector are more likely to live in non-rural areas compared to workers in the informal sector, workers in the two sectors are not geographically segregated and they all face the same labour market environment at least at the province level. Lastly, IFLS is known for its comprehensive data collection range. To be more specific, IFLS collects information regarding age, education, marital status and residence location, all of which not only affect whether a person works in the formal sector or the informal sector but also have a direct impact on our outcome of interests, i.e. earnings.

The other key factor determining whether the matching estimates are reliable is the matching method itself. For instance, pair-wise matching estimates, especially closest-neighbour matching, though better than OLS estimates, in many cases cannot duplicate the experimental estimates even with high-quality data satisfying the above three conditions. On the other hand, these papers also find that difference-in-difference conditioned on matched propensity scores is the matching estimator that has best overall performance in most cases. In our study, we report local-linear-regression matching estimates with bootstrap.

III. Data

i. IFLS

The main analysis of this paper relies on the second and the third wave of the Indonesian Family Life Survey (IFLS), a longitudinal survey on the individual, household, and community level. Thanks to its extremely low attrition rate Strauss et al. (2016), IFLS allows us to construct labour market participation history across years and across waves. Another advantage of IFLS is that it documents data on the self-employed individuals, enabling us to analyse the dominant employment sector in a typical developing country. While we use annual wage income as the income for private workers and government workers, we use annual net profit as income for the self-employed. We drop all the observations with negative net profit, considering that net profit takes out all business expenses including returns to capital. In addition, we drop all observations who works as unpaid family workers.

Among five waves of IFLS, we conduct our study using the 1997 and 2000 wave due to the availability of firm size data and data on income in the last year. First, firm size is an influential factor in determining the classification of the formal and informal sector. However, not until the second wave, the team includes “number of people in the work place" in the questionnaires regarding the primary job in the survey year. Second, since the income in the last year is significantly correlated with the current income and the current employment sector, it is a covariate so crucial to be omitted if we were to control for the selection. Unfortunately, only the first three waves of IFLS collect income data on jobs in between waves.[3]

The data set in 1997 include 4,270 observations of adult male of working age (between 23 and 60 years) who report to be employed or mention “work" as their primary activity. On the other hand, in 2000 we have 6,425 such observations. Note that the two samples only have partial overlap: people who work in 1997 may retire or exit the labour market; meanwhile, those who appear in 2000 but not in 1997 can either be young people who enter the labour force after 1997 or new household members who join the target household as the spouse of the target individuals.[4]

[3] But we still find significant wage premium even when we do not use firm size in the definition.

[4] The latter is also the reason why there are so many more observations in 2000: young people start up their own family and their spouses in the split-off family become target individuals.

ii. The Classification of The Formal and Informal Sector

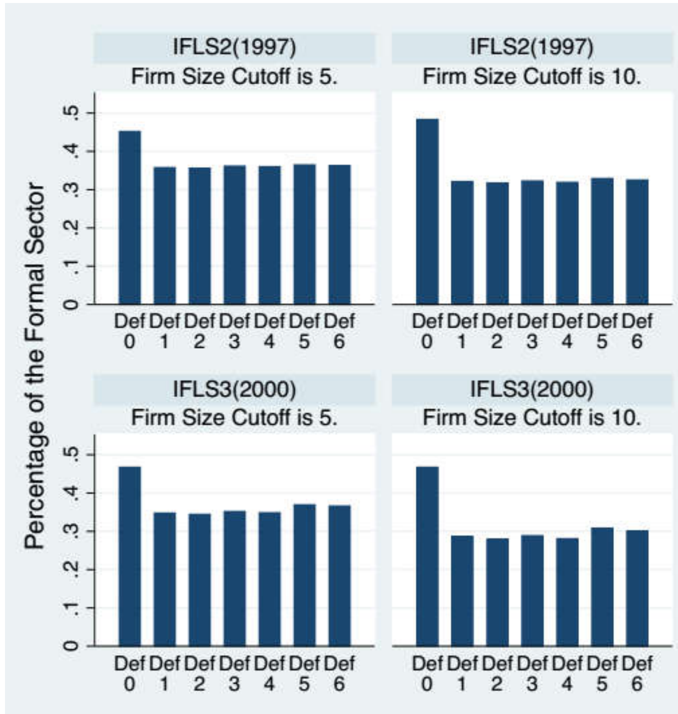

Figure 1: *Percentage of The Formal Sector Under All Definitions Using Different Firm Size Threshold*

Our study provides thirteen definitions of the formal and informal sector. In the appendix, we present the classification rules of all the definitions and explain why we are interested in these different classifications.

Compared to other studies using IFLS, our study goes great length to refine the classification of the two sectors. Many IFLS studies, such as Hohberg and Lay (2015), define the sector of employment solely in accordance to the self-reported working status: they classify workers as formal that are either “private" or “government workers" and as informal those workers who report “self-employed" or “unpaid family worker" as their working status. In contrast, even the crudest classification in this study can be regarded as a refinement of their work-status classification using occupation types. Compared to the work-status classification, the benchmark classification (hereafter coded as “Definition 0") of this study makes the following two major changes in the classification: first, professional workers in comparison to general workers, agricultural workers and managers are classified as formal workers regardless of their working status; second, private workers whose primary duties are non-managerial agricultural-related jobs are classified as informal workers.

Note that our bench mark classification regards all government workers, professionals and technicians, private workers whose primary duties are supervisors and managers and private workers who are general workers as formal workers. Furthermore, we use firm size to further refine our classification of the two sector: only people whose firm size is above certain threshold are regarded formal workers. We apply the firm size rule to different types of workers as robustness

Table 1: *Summary Statistics(1997): Def I with 5 as firm size cutoff point*

Variables	Informal	Formal	MeanDiff
income in last year	10.84	11.47	-0.626***
age	40.36	37.47	2.896***
education			
own	5.575	9.639	-4.064***
father's	2.271	4.349	-2.078***
mother's	1.466	3.070	-1.604***
urban	0.346	0.695	-0.349***

Table 2: *Summary Statistics(2000): Def I with 5 as firm size cutoff point*

Variables	Informal	Formal	MeanDiff
income in last year	10.81	11.27	-0.452***
age	39.05	36.69	2.365***
education			
own	6.528	10.55	-4.020***
father's	0.464	0.984	-0.520***
mother's	0.294	0.657	-0.363***
urban	0.403	0.690	-0.287***

check. There are two firm size cut-off points in our analysis: five and ten. For each firm size cut-off point, we have six definitions. Therefore, there are twelve definitions using firm size information together with the benchmark definition.

Figure 1 shows the percentage of the formal sector under all thirteen definitions in 1997 and 2000. It shows that applying firm size rule to the classification of the two sectors reduces the size of the formal sector drastically. Unsurprisingly, we also find that the bigger the firm size cut-off point is, the smaller the formal sector is. However, according to the Figure 1, it seems that it does not matter to which types of workers we apply the firm size rule to. When using the same firm size threshold, the size of formal sector does not vary much under all six definitions using firm size, namely from Definition 1 to Definition 6.

iii. Evidence of Selection

Before examining the effects of being formally employed on earnings, we need to identify whether one of the employment sectors attracts more competent workers. In other words, are workers with higher productivity hired in the formal sector and those who are less competitive hired in the informal sector? Though in our dataset, we do not observe ability and productivity directly, we do observe education and other individual characteristics that are correlated with ability. If we do observe that informal workers and formal workers are fundamentally different from each other in terms of the observed characteristics, it is reasonable for us to assume that workers in the two sectors differ from each other in terms of unobserved characteristics as well.

For the ease of explanation, among the thirteen definitions of the formal sector, here we only present the evidence for the cream-skimming effect in the formal sector observed under Definition 1 when using five as the firm size threshold in both 1997 and 2000. [5]

[5] The results for the comparison of workers' characteristics in the two sectors under the other definitions, including benchmark definition, are included in the appendix.

As shown in Table 1 and Table 2, workers in the two sectors differ significantly in the perspective of five types of observed characteristics: their income in last year, age, own education, parents' education and residence location. All five types of characteristics have been widely accepted as factors that have impacts on earnings.

The universal pattern we observed in 1997 and 2000 is that workers in the formal sector are significantly better-off than informal workers in all five perspectives. First, they tend to be younger, have more years of schooling, and their parents are better-educated than the parents of informal workers. Second, formal workers are more likely to reside in urban areas than informal workers. Last but not the least, regardless of their employment sector in the last year, formal workers earn more than the informal worker last year.

To provide additional evidence of selection, we provide the Logit estimates for being in the formal sector. Again, for ease of explanation, we only discuss the propensity score estimates when using five and more workers as the firm size threshold in the classification of the formal and informal sector,though in appendix we do present the Logit estimates under all other definitions using ten as the firm size threshold. Table 5 and Table 7 show the Logit estimates of being in the formal sector under Definition 1 to Definition 6 when using five as the threshold in 1997 and 2000 as well as under Definition 0 which does not use firm size to classify workers. We also include province dummies to capture provincial fixed effects.

From the two tables, we can conclude the following regarding the selection. First, income in the last year is a good predicator of current employment sector. The more one earns in the last year, the more likely he works in the formal sector the current year. Second, higher education level increases the probability of working in the formal sector, though the impact of parent's education is not as influential. Third, though the impact of age on the sector is not significant in 1997 sample, in 2000 age has an inverted U-shaped impact on the likelihood of working in the formal sector. Last but not the least, as one would expect, living in urban area in the last year increase the likelihood of finding a formal job.

IV. Econometric Method

Following (Dehejia and Wahba, 1999, 2002; Ham et al., 2016; Heckman et al., 1997, 1998; Rosenbaum and Rubin, 1983), we calculate the average treatment effect, i.e. the measurement of the formal earning sector premium, using the following equation:

$$\alpha_{ATE} = E[Y^{F=1} - Y^{F=0}], \tag{1}$$

where F denotes the treatment status of the individual, taking the value 1 if we let the individual work in the formal sector, and 0 if we let him work in the informal sector; Y denotes the outcome of interest, namely the earnings of the individual and the superscript indicates his or her earnings in different sectors: $Y^{F=1}$ is how much the individual can earn if he or she works in the formal sector and $Y^{F=0}$ is the amount he can earn if he works in the informal sector.

Note that α_{ATE} is not equivalent to the simple difference in the average earnings between formal workers and informal workers. In effect, the simple difference can be a trust-worthy measurement for the formal earning premium if and only if the treatment is randomly assigned to the population. Unfortunately, there is well-established evidence in the literature supporting selection between the formal sector and the informal sector. As mentioned in the previous section, we find observable differences between workers in the formal and informal sector, which indicating that more likely than not workers in the two sectors are not comparable in unobservables that are correlated with earnings capabilities.

Consider worker i, who could either work in the formal sector or the informal sector, and an experiment switching i from the informal sector to the formal sector. The change in earnings is ATE.

Recall that worker i could earn $Y_i^{F=1}$ if he works in the formal sector and $Y_i^{F=0}$ if in the informal sector. When worker i is a formal sector, i.e. $Sector_i = 1$, the difference between his earnings when he works in the formal sector and his earnings when he works in the informal sector is ATT. For ease of exposition, we will explain how to get ATT in the following part of this chapter and ATU is calculated in the same manner with the only difference being that in ATU, i is an informal worker.

As shown below, ATT is defined as the average earning differences between the formal earnings and the informal earnings of formal workers:

$$\begin{aligned} \alpha_{ATT} &= E[(Y^{F=1} - Y^{F=0})|Sector = 1)] \\ &= E[Y^{F=1}|Sector = 1] - E[Y^{F=0}|Sector = 1] \end{aligned} \tag{2}$$

Note that in contrast to the conceptual treatment status variable F, the variable $Sector$ is the sector where individual i works in his real life. Therefore, we cannot observe the counter-factual average of earnings in the informal sector for those are in the formal sector, i.e the second term in the above equation.

Due to the selection bias, we cannot use $E(Y^{F=0}|Sector = 0)$ as the counterpart for $E(Y^{F=0}|Sector = 1)$. The propensity score matching method provides us a solution when the following two assumptions are satisfied.

1. Conditional Independence Assumption (i.e. CIA) for ATT: we have X such that

$$Y^{F=0} \perp Sector|X, \tag{3}$$

where X is a vector of observed pre-treatment characteristics. This assumption states that conditional on X, the assignment of treatment, i.e. $Sector$ is random with respect to how much the worker can earn in the informal sector. Hence, all us to use $E[Y^{F=0}|X, Sector = 0]$ as the counter-factual $E[Y^{F=0}|X, Sector = 1]$.

Essentially, CIA allows us to use the outcome of the untreated who are similar to the untreated in respect to observed pre-treatment characteristics to proximate for the outcome of a treated individual.

Note that, X is a vector of pre-treatment characteristics. If the dimension of X is big enough, we might not be able to find an untreated individual with the same value of X as for a treated individual. Hence, we match individuals who are similar in terms of the propensity score, namely the probability of getting treatment conditional on X:

$$\hat{Pr}(F = 1|X) = \frac{1}{1 + e^{-X\hat{\beta}}}, \tag{4}$$

where we take logistic form for the probability of being treated here and $\hat{\beta}$ are the estimated coefficient of the logit function.

2. Common Support Assumption for ATT:

$$0 \leq Pr(F = 1|X, Sector = 1) < 1. \tag{5}$$

This assumption guarantees that we can find untreated individuals who are with the same propensity score as the treated. Note that no one in the untreated group could have propensity score equals to 1, since $P(X) = 1$ suggests that this individual will always get the treatment. Therefore, there will be no match for formal workers whose probability of being treated is 1 in the untreated group. However, we do not usually observe people with $P(X) = 1$ or $P(X) = 0$ in the data. Therefore, the common support assumption for ATT requires us just to drop from the sample those with $P(X) \to 1$.

Unlike Heckman et al. (1997), in which including the effect on people who are not supposed to be treated makes ATE less appealing to the policy makers than ATT, we are interested in the treatment effect on the informal workers as well. Hence, we calculated the average treatment effect on the untreated as well, using the equation below:

$$\begin{aligned}\alpha_{ATU} &= E[(Y^{F=1} - Y^{F=0})|Sector = 0)] \\ &= E[Y^{F=1}|Sector = 0] - E[Y^{F=0}|Sector = 0]. \end{aligned} \tag{6}$$

α_{ATU} explicitly evaluates the effect of working in the formal sector on those who are in the informal sector.

Similarly, to make sure that we can use the propensity score matching method to calculate α_{ATU}, we need to make the following two assumptions.

1. Conditional Independence Assumption (i.e. CIA) for ATU:

$$Y^{F=1} \perp Sector|X, \tag{7}$$

where X is a vector of observed pre-treatment characteristics. This assumption makes sure that conditional on X, the assignment of treatment, i.e *Sector* is random with respect to how much the worker can earn in the formal sector. Hence, all us to use $E[Y^{F=1}|X, Sector = 1]$ as the counter-factual $E[Y^{F=1}|X, Sector = 0]$.

2. Common Support Assumption for ATU:

$$0 < Pr(F = 1|X, Sector = 1) \leq 1. \tag{8}$$

This assumption guarantees that we can find treated individuals who are with the same propensity score as the untreated. Note that no one in the treated group could have propensity score equals to 0, since $P(X) = 0$ suggests that this individual will never get the treatment. Therefore, there will be no match for informal workers whose probability of being treated is 0 in the treated group. As with ATT, since we hardly observe anyone in the data with $P(X) = 0$, we drop untreated individuals with $P(X) \to 0$ instead.

As mentioned, ATE informs us of the difference between i's earnings in the formal sector and his earnings in the informal sector, provided we randomly choose i from the population. Therefore, we could calculate ATE, using the following equation:

$$\begin{aligned}\alpha_{ATE} &= E_{Sector}[(Y^{F=1} - Y^{F=0})|Sector] \\ &= P(Sector = 0) \times \Big\{E[Y^{F=1}|Sector = 0] - E[Y^{F=0}|Sector = 0]\Big\} \\ &+ P(Sector = 1) \times \Big\{E[Y^{F=1}|Sector = 0] - E[Y^{F=0}|Sector = 1]\Big\}, \end{aligned} \tag{9}$$

where $P(Sector = 0)$ and $P(Sector = 1)$ are the percentage of informal workers and formal workers in the real labour market.

For ATE to be valid, we need to assume that the CIA and Common Support Assumptions for both ATT and ATU are satisfied. That is, equivalent to the following two assumptions.

1. Conditional Independence Assumption (i.e. CIA) for ATE:

$$(Y^{F=0}, Y^{F=1}) \perp Sector|X. \tag{10}$$

2. Common Support Assumption for ATU:

$$0 < \hat{Pr}(F = 1|X, Sector = 1) < 1. \tag{11}$$

The X in our study includes age, education, one-year-lagged location type and one-year-lagged earnings. All the earnings are in real terms and we use the natural logarithm of the real earnings.

V. Results

As before, we only discuss the results using five as the firm size threshold in the classification. [6] Table 33 and Table 35 show the estimates of wage premium using 1997 and 2000 data. The first three rows correspond to the three measures of wage premium using matching method. On the other hand, we also report the ordinary least square estimates of wage premium when controlling for individual characteristics in linear form in the OLS row. In addition, we provide the mean difference of income between the two sectors as a reference point.

Compared with the mean difference estimates, both OLS estimates and the three measures of wage premium using propensity score matching reduce the wage gap between the two sectors substantially. Therefore, we know that the observed wage gap in the average income between the two sectors are returns to younger age, higher education. better residence location and higher ability (reflected as higher income in the last year).

Moreover, note that compared to the wage premium under the benchmark definition of the formal sector (Definition 0), the estimates under the definitions incorporated firm size (Definition 1 to Definition 6) are clearly smaller. Hence, it is really important to use the firm size information in the definition.

Meanwhile, if we compare the OLS estimates to the matching estimates, we cannot see any significant difference. This may suggest that though the selection is too important to be ignored when measuring the wage premium, how we control for the selection is not as crucial. Simple OLS seems work as good as matching.

To see whether propensity score matching reduces the extent of selection between the two sectors, we run balance test on the matched sample. After trimming samples off the common support, we use the same estimated propensity scores to check whether there is still treatment effect on the explanatory variables in the propensity functions. If the matching method successfully balances the sample, there should be no treatment effect. Table 3 and 4 report the number of unbalanced variables under all nineteen definitions. Each cell informs us the number of unbalanced variables when the significance level is 10%, 5% and 1%.

Table 3: *IFLS2: Balance Tests for All Explanatory Variables in The Propensity Functions*

	Def 0	Def 1	Def 2	Def 3	Def 4	Def 5	Def 6
A. Firm Size Threshold is 5.							
ATT	1/1/1	0/0/2	0/1/3	0/0/1	0/0/1	0/0/0	0/0/1
ATU	0/2/3	1/1/2	1/1/2	0/1/2	0/1/2	0/1/1	0/1/1
ATE	0/0/0	0/1/1	0/1/1	0/0/1	0/0/1	0/0/1	0/1/1
% of The Formal Sector	45.18%	35.76%	35.60%	36.18%	36.02%	36.49%	36.32%
B. Firm Size Threshold is 10.							
ATT	1/1/1	0/1/1	0/1/1	0/1/1	0/1/1	0/0/0	0/0/0
ATU	0/2/3	2/2/2	2/2/2	2/2/2	2/2/2	2/2/2	2/2/2
ATE	0/0/0	2/2/3	2/2/3	2/2/3	2/2/3	2/2/2	2/2/2
% of The Formal Sector	45.18%	30.05%	29.70%	30.19%	29.84%	30.77%	30.42%

Note: Total number of explanatory variables tested is 19 since we exclude the provincial dummies from the test.

Based on the two tables, we have the following findings. First, for both 1997 and 2000 sample, under all definitions, we can get the most optimistic balancing results when using ATT as the measure whether the explanatory variable passes the balance test compared to ATU and ATE.

[6] We do provides robustness check using ten as the threshold in the appendix.

Table 4: *IFLS3: Balance Tests for All Explanatory Variables in The Propensity Functions*

	Def 0	Def 1	Def 2	Def 3	Def 4	Def 5	Def 6
A. Firm Size Threshold is 5.							
ATT	0/1/1	0/0/2	0/0/2	0/2/2	0/2/2	0/2/2	0/2/2
ATU	0/3/3	1/3/4	1/3/3	0/2/4	1/3/4	0/3/3	1/3/3
ATE	0/1/2	0/1/1	0/1/2	0/1/1	0/1/1	0/1/1	0/1/2
% of The Formal Sector	46.74%	34.77%	34.44%	35.15%	34.85%	36.93%	36.61%
B. Firm Size Threshold is 10.							
ATT	0/1/1	0/2/2	0/2/2	0/2/2	0/2/2	2/2/2	2/2/2
ATU	0/3/3	2/2/5	2/2/5	2/2/5	2/2/4	2/2/5	2/4/4
ATE	0/1/2	1/2/2	2/2/2	2/2/2	2/2/2	0/2/2	0/2/2
% of The Formal Sector	46.74%	28.68%	27.91%	28.86%	28.08%	30.85%	30.07%

Note: Total number of explanatory variables tested is 19 since we exclude the provincial dummies from the test.

One possible reason is that the untreated individuals, i.e. the informal workers, account for more than 50% of the full sample. Therefore, when calculating ATU, we do not have enough treated individuals(formal workers) to be matched to the informal workers. In effect, this is the very reason why the sample under Definition 0 is more likely to passes the balance test than the other definitions with firm size threshold. As we can see, the percentage of the formal sector is closer to 50% under Definition 0 than Definition 1 to Definition 6. Next, we can see that according to the ATT measure, matching method is more successful for samples using higher firm size threshold.

In sum, we can conclude that at the 5% singificance level, according to the ATT estimates, propensity score matching method successfully constructs balanced sample.

VI. Discussion

A. Definitions Of The Formal Sector

i. Data Source

The key variable in our study is F, the employment sector. In IFLS questionnaire, we have three questions that can help us identify which sector the interviewee belongs to. First, the work status of his job. Second, the description of his primary duties. Third, the number of people in his work place. We will explain why we believe the answers to the three questions can help with the classification.

i.1 Work Status

For those who are working, when being asked about their work status, they can choose from the following answers: self-employed, private workers, government workers, unpaid family workers and casual workers.[7] We exclude unpaid family workers from our sample, though they seem to be a natural part of the informal sector. Since they will not be asked how much is their income for their unpaid family worker job, including them in the informal sector will only produce bigger wage differences between the formal and informal sector.

i.2 Occupation Codes

The IFLS team asks the interviewees to describe their primary duties. And according to the answers, the interviewers pick one of the one hundred occupation codes which are the best-fitted for that description. We further classify the one hundred occupations into four types: first, professional and governmental jobs such as professors, doctors, lawyers and government officials; second, managerial jobs such as store managers and clerical supervisors; third, non-managerial jobs in agriculture, such as planters, fishermen and hunters; last, all other workers, namely general workers who are not in agriculture, such as production workers and service workers.

i.3 Firm Size

Common senses tell us that the bigger the firm is, the less likely it can evade the government supervision. Therefore, besides working status and occupation types, we exploit data on “number of people in the workplace" to further refine our classification of the formal and informal sector. The reported number of people in the workplace can be an approximate value for the firm size. When we apply firm size criteria to certain types of workers, they will be classified as formal workers if and only if the firm size is above the threshold we choose.

Considering that Indonesia’s Central Bureau of Statistics (BPS) define micro and small firms using five and twenty as the threshold, to make sure our results are insensitive to the choice of firm size threshold, we report our results using both five and twenty as the threshold. Furthermore, since the 80-percentile firm size in our IFLS sample is eight workers, we also provide results using ten as the threshold.

[7] In every wave of IFLS, except IFLS 2, there are three types of self-employed workers: self-employed with no workers, self-employed with temporary workers or unpaid family workers, and self-employed with paid workers. However, since IFLS 2 does not differentiate those three types of self-employed workers, we regard them as the same. In addition, since the category of casual workers is only introduced in IFLS 4 and IFLS 5, those who are casual workers could either report themselves as private workers or the self-employed. Fortunately, as mentioned since IFLS 4 and IFLS 5 do not have lagged income which is an important predictor of current employment sector, we only use IFLS 2 and IFLS 3 in our sample. Hence, we do not need to worry about this problem.

ii. Classification Rules

In this section, we will explain how we use the above variables to classify the formal and informal sector.

ii.1 Benchmark Definition

In our benchmark definition, we use only work status and occupation codes. And the definition rules are as following:

- *The Formal Sector* includes:
  1. Professionals and government workers;
  2. Private workers whose occupations are managerial jobs or non-agricultural general jobs;
- *The Informal Sector* includes:
  1. Private workers who are doing non-managerial jobs in agriculture;
  2. Self-employed whose primary duties are not professionals.

In the benchmark definitions, most of the self-employed are classified as informal workers and most of the private workers formal workers. The two major exceptions are: first, we classify professionals and government workers as formal workers regardless of their working status;[8] second, agricultural workers who are hired by private entities are classified as informal jobs.

ii.2 Definitions Apply Firm Size Rules

The most important exception in the benchmark definition is the exception regarding the professionals. In the other definitions using firm sizes, those workers are regarded as formal workers regard less of the firm size. The main argument for this exception is that for those professionals, the nature of their jobs do not require economics of scale: a lawyer who hires one or two assistant and opens his own start-up firm can be as formal as lawyers in large law firms.

However, one might feel suspicious that we believe workers with the same working status and occupation types should always be in the same sector. In effect, we do observe counterexample in life: both waiters in McDonald's and cook in a family-owned small restaurant private fall into the category of general workers who are private workers. However, we tend to regard the former as formal workers and the latter informal workers. Because we believe the bigger firms are more likely to offer formal employment than the smaller firms which can evade governmental supervision more easily.

Unlike the classification private general workers, there is ambiguity regarding the classification of the following the two types of workers: i) self-employed managers, and ii) private workers whose primary duties are supervisors and managers. We can either regard self-employed managers as formal workers, informal workers or apply the firm size threshold rule. Consequently, we have another six classifications, which we denote as Definition 1 to Definition 6. Definition 1 and Definition 2 regard self-employed managers as informal workers; Definition 3 and Definition 4 regard self-employed managers as informal workers if and only if there are strictly less than certain amount of people (i.e. depending on the firm size threshold we choose) in their work places; Definition 5 and Definition 6 classify self-employed managers into the formal sector as

[8]For instance, both a self-employed doctor and a doctor who works in a private hospital are regarded as formal workers.

the benchmark classification does. On the other hand, since bigger firms are more likely to hire professional managers (private workers whose primary duties are managerial) while owners of smaller firms can play the role of managers themselves (self-employed managers), we either regard private managers as formal workers (in the odd classifications) or apply the firm size threshold rules to them (in the even classifications). In total, we have nineteen different definitions of the formal sector including the definition of the formal sector which is based on work status and occupation codes. Figure 2 shows the differences among Definition 1 to Definition 6.

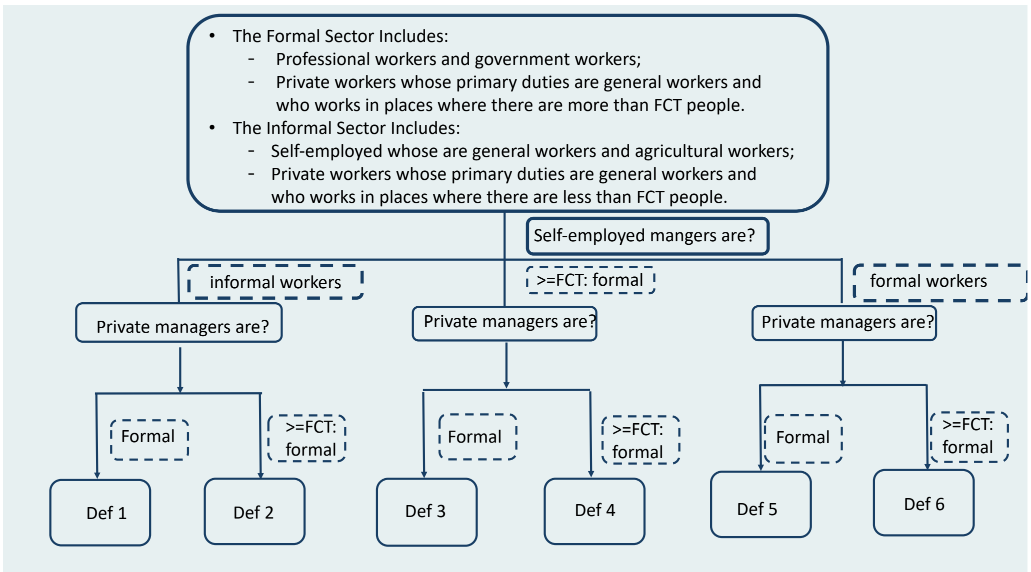

Figure 2: *Comparison of Definitions Using Firm Size*

From Figure 2, we can see that all six definitions have reach an agreement regarding how to classify the following three groups of workers: first, they all regard professional workers and government workers as formal workers; second, they classify the self-employed whose are general workers or agricultural workers as informal workers; last but not the least, all six definitions apply firm size rule to general workers who are hired by private firms, namely only those workers whose workplace have FCT or more people, where FCT is the firm size threshold of our choice.

On the other hand, Figure 2 also shows that the main difference among the six definitions are the following two aspects: first, whether the self-employed managers are regarded as formal or informal workers or even we need the help of firm size to determine which sector they belong to; second, whether all managers and supervisors who are hired by private firms should be regarded as formal workers or only those who work in places with more than FCT people should be classified as formal workers.

Table 5: *IFLS2(1997): Firm Size Cutoff is 5*

	(1) Def 0	(2) Def 1	(3) Def 2	(4) Def 3	(5) Def 4	(6) Def 5	(7) Def 6
Last year's income	0.237***	0.243***	0.238***	0.274***	0.269***	0.283***	0.278***
	(0.0415)	(0.0443)	(0.0443)	(0.0446)	(0.0446)	(0.0447)	(0.0447)
Age	-0.0182	0.0272	0.0282	0.0338	0.0347	0.0301	0.0311
	(0.0315)	(0.0331)	(0.0330)	(0.0332)	(0.0332)	(0.0333)	(0.0332)
Age Square	-0.000191	-0.000595	-0.000602	-0.000682	-0.000689	-0.000623	-0.000630
	(0.000386)	(0.000408)	(0.000407)	(0.000410)	(0.000409)	(0.000410)	(0.000409)
Own Education IS							
Primary school graduates or some junior high school	0.421***	0.351***	0.368***	0.331**	0.349***	0.330**	0.347***
	(0.0944)	(0.106)	(0.106)	(0.106)	(0.106)	(0.105)	(0.105)
Junior high school graduates or some senior high school	0.728***	0.775***	0.785***	0.728***	0.738***	0.731***	0.740***
	(0.125)	(0.131)	(0.131)	(0.131)	(0.132)	(0.131)	(0.131)
Senior high school graduates	1.399***	1.711***	1.715***	1.702***	1.705***	1.714***	1.716***
	(0.113)	(0.117)	(0.117)	(0.117)	(0.117)	(0.117)	(0.117)
Some college or above	1.763***	2.234***	2.260***	2.302***	2.328***	2.436***	2.462***
	(0.243)	(0.237)	(0.237)	(0.244)	(0.244)	(0.252)	(0.252)
Father's Education IS							
Primary school graduates or some junior high school	0.0234	0.0591	0.0599	0.0241	0.0250	0.0627	0.0636
	(0.103)	(0.106)	(0.106)	(0.106)	(0.106)	(0.106)	(0.106)
Junior high school graduates or some senior high school	0.0817	0.139	0.154	0.216	0.232	0.244	0.259
	(0.187)	(0.181)	(0.181)	(0.183)	(0.183)	(0.184)	(0.184)
Senior high school graduates	0.0214	0.100	0.0587	0.106	0.0642	0.107	0.0648
	(0.230)	(0.221)	(0.220)	(0.223)	(0.222)	(0.224)	(0.223)
Some college or above	0.289	0.246	0.262	0.782	0.795	1.649	1.659
	(0.813)	(0.714)	(0.712)	(0.830)	(0.829)	(1.097)	(1.095)
Mother's Education IS							
Primary school graduates or some junior high school	0.154	0.0845	0.0659	0.119	0.100	0.0922	0.0731
	(0.114)	(0.115)	(0.115)	(0.115)	(0.115)	(0.116)	(0.116)
Junior high school graduates or some senior high school	0.156	-0.111	-0.0973	-0.170	-0.155	-0.248	-0.233
	(0.259)	(0.241)	(0.241)	(0.244)	(0.243)	(0.245)	(0.245)
Senior high school graduates or above	-0.0825	-0.0871	-0.114	-0.187	-0.213	-0.249	-0.275
	(0.313)	(0.294)	(0.292)	(0.296)	(0.294)	(0.298)	(0.296)
Urban Area	1.040***	0.914***	0.911***	0.940***	0.937***	0.950***	0.946***
	(0.0784)	(0.0822)	(0.0822)	(0.0824)	(0.0824)	(0.0825)	(0.0824)
Constant	-3.570***	-5.028***	-4.996***	-5.485***	-5.451***	-5.529***	-5.495***
	(0.745)	(0.781)	(0.780)	(0.786)	(0.785)	(0.787)	(0.785)
Province Dummy	yes	yes	yes	yes	yes	yes	yes
N	4270	4270	4270	4270	4270	4270	4270

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: *IFLS2(1997): Firm Size Cutoff is 10*

	(1) Def 0	(2) Def 1	(3) Def 2	(4) Def 3	(5) Def 4	(6) Def 5	(7) Def 6
Last year's income	0.237***	0.252***	0.251***	0.268***	0.266***	0.296***	0.294***
	(0.0415)	(0.0469)	(0.0470)	(0.0471)	(0.0471)	(0.0472)	(0.0472)
Age	-0.0182	0.0812*	0.0861*	0.0868*	0.0917**	0.0844*	0.0893*
	(0.0315)	(0.0348)	(0.0349)	(0.0349)	(0.0350)	(0.0349)	(0.0350)
Age Square	-0.000191	-0.00109*	-0.00115**	-0.00116**	-0.00122**	-0.00112**	-0.00118**
	(0.000386)	(0.000429)	(0.000430)	(0.000431)	(0.000431)	(0.000430)	(0.000431)
Own Education IS							
Primary school graduates or some junior high school	0.421***	0.457***	0.473***	0.452***	0.467***	0.430***	0.446***
	(0.0944)	(0.118)	(0.119)	(0.118)	(0.119)	(0.117)	(0.118)
Junior high school graduates or some senior high school	0.728***	1.039***	1.065***	1.024***	1.050***	0.986***	1.012***
	(0.125)	(0.141)	(0.141)	(0.141)	(0.141)	(0.140)	(0.140)
Senior high school graduates	1.399***	1.988***	1.996***	1.979***	1.988***	1.973***	1.981***
	(0.113)	(0.125)	(0.125)	(0.125)	(0.125)	(0.124)	(0.124)
Some college or above	1.763***	2.547***	2.555***	2.596***	2.601***	2.701***	2.700***
	(0.243)	(0.232)	(0.230)	(0.235)	(0.233)	(0.242)	(0.240)
Father's Education IS							
Primary school graduates or some junior high school	0.0234	-0.0172	-0.0385	-0.0323	-0.0537	-0.0134	-0.0349
	(0.103)	(0.110)	(0.110)	(0.110)	(0.110)	(0.110)	(0.110)
Junior high school graduates or some senior high school	0.0817	-0.0204	-0.0304	0.00822	-0.00203	0.0760	0.0654
	(0.187)	(0.180)	(0.180)	(0.180)	(0.180)	(0.181)	(0.181)
Senior high school graduates	0.0214	-0.0836	-0.117	-0.0955	-0.128	-0.0793	-0.112
	(0.230)	(0.216)	(0.215)	(0.216)	(0.216)	(0.218)	(0.217)
Some college or above	0.289	-0.0814	-0.417	0.329	-0.0775	0.879	0.340
	(0.813)	(0.647)	(0.606)	(0.711)	(0.646)	(0.826)	(0.713)
Mother's Education IS							
Primary school graduates or some junior high school	0.154	0.0630	0.0567	0.0693	0.0632	0.0700	0.0642
	(0.114)	(0.117)	(0.117)	(0.117)	(0.117)	(0.118)	(0.118)
Junior high school graduates or some senior high school	0.156	0.160	0.208	0.158	0.208	0.0402	0.0934
	(0.259)	(0.237)	(0.237)	(0.239)	(0.238)	(0.240)	(0.240)
Senior high school graduates or above	-0.0825	0.211	0.167	0.171	0.126	0.0600	0.0136
	(0.313)	(0.285)	(0.281)	(0.285)	(0.281)	(0.287)	(0.283)
Urban Area	1.040***	0.765***	0.749***	0.776***	0.760***	0.801***	0.785***
	(0.0784)	(0.0870)	(0.0872)	(0.0870)	(0.0872)	(0.0870)	(0.0871)
Constant	-3.570***	-6.757***	-6.856***	-7.015***	-7.113***	-7.299***	-7.393***
	(0.745)	(0.821)	(0.822)	(0.824)	(0.825)	(0.826)	(0.826)
Province Dummy	yes	yes	yes	yes	yes	yes	yes
N	4270	4270	4270	4270	4270	4270	4270

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: *IFLS2(2000): Firm Size Cutoff is 5*

	(1) Def 0	(2) Def 1	(3) Def 2	(4) Def 3	(5) Def 4	(6) Def 5	(7) Def 6
Last year's income	0.0912**	0.109**	0.112***	0.138***	0.142***	0.163***	0.167***
	(0.0311)	(0.0336)	(0.0336)	(0.0337)	(0.0337)	(0.0333)	(0.0333)
Age	0.0410	0.0960***	0.0975***	0.0957***	0.0972***	0.0921***	0.0934***
	(0.0241)	(0.0259)	(0.0259)	(0.0259)	(0.0259)	(0.0255)	(0.0255)
Age Square	-0.000822**	-0.00125***	-0.00128***	-0.00125***	-0.00127***	-0.00118***	-0.00120***
	(0.000300)	(0.000323)	(0.000324)	(0.000323)	(0.000324)	(0.000318)	(0.000318)
Own Education IS							
Primary school graduates or some junior high school	0.440***	0.537***	0.522***	0.524***	0.508***	0.492***	0.478***
	(0.0816)	(0.0967)	(0.0968)	(0.0964)	(0.0965)	(0.0934)	(0.0935)
Junior high school graduates or some senior high school	0.696***	0.879***	0.868***	0.850***	0.839***	0.794***	0.784***
	(0.0973)	(0.109)	(0.110)	(0.109)	(0.109)	(0.107)	(0.107)
Senior high school graduates	1.333***	1.847***	1.817***	1.840***	1.810***	1.805***	1.774***
	(0.0915)	(0.101)	(0.101)	(0.100)	(0.100)	(0.0982)	(0.0982)
Some college or above	1.717***	2.273***	2.270***	2.268***	2.265***	2.236***	2.232***
	(0.160)	(0.158)	(0.158)	(0.159)	(0.159)	(0.160)	(0.160)
Father's Education IS							
Primary school graduates or some junior high school	-0.0200	0.133	0.139	0.147	0.152	0.186*	0.192*
	(0.0761)	(0.0792)	(0.0792)	(0.0792)	(0.0792)	(0.0785)	(0.0785)
Junior high school graduates or some senior high school	0.267*	0.392**	0.388**	0.409**	0.405**	0.348**	0.344**
	(0.132)	(0.128)	(0.128)	(0.129)	(0.129)	(0.129)	(0.129)
Senior high school graduates	0.00560	0.0997	0.108	0.0903	0.0991	0.112	0.121
	(0.148)	(0.143)	(0.143)	(0.144)	(0.144)	(0.145)	(0.144)
Some college or above	0.159	0.174	0.249	0.197	0.273	0.0997	0.178
	(0.393)	(0.352)	(0.352)	(0.359)	(0.359)	(0.359)	(0.359)
Mother's Education IS							
Primary school graduates or some junior high school	0.209*	0.121	0.122	0.115	0.116	0.106	0.108
	(0.0833)	(0.0845)	(0.0844)	(0.0846)	(0.0845)	(0.0843)	(0.0842)
Junior high school graduates or some senior high school	0.0240	0.0102	0.0121	0.0288	0.0306	0.0654	0.0669
	(0.163)	(0.157)	(0.157)	(0.158)	(0.158)	(0.159)	(0.159)
Senior high school graduates or above	-0.298	-0.169	-0.245	-0.116	-0.194	-0.122	-0.201
	(0.181)	(0.174)	(0.173)	(0.175)	(0.174)	(0.176)	(0.175)
Urban Area	0.923***	0.650***	0.633***	0.654***	0.637***	0.679***	0.662***
	(0.0627)	(0.0669)	(0.0669)	(0.0669)	(0.0669)	(0.0661)	(0.0660)
Constant	-3.347***	-5.606***	-5.656***	-5.941***	-5.989***	-6.057***	-6.100***
	(0.553)	(0.593)	(0.594)	(0.595)	(0.595)	(0.588)	(0.588)
Province Dummy	yes	yes	yes	yes	yes	yes	yes
N	6424	6424	6424	6424	6424	6424	6424

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8: *IFLS2(2000): Firm Size Cutoff is 10*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Def 0	Def 1	Def 2	Def 3	Def 4	Def 5	Def 6
Last year's income	0.0912**	0.131***	0.128***	0.150***	0.147***	0.191***	0.188***
	(0.0311)	(0.0362)	(0.0363)	(0.0362)	(0.0363)	(0.0355)	(0.0356)
Age	0.0410	0.113***	0.116***	0.115***	0.118***	0.107***	0.110***
	(0.0241)	(0.0278)	(0.0279)	(0.0278)	(0.0279)	(0.0271)	(0.0272)
Age Square	-0.000822**	-0.00134***	-0.00137***	-0.00136***	-0.00139***	-0.00124***	-0.00127***
	(0.000300)	(0.000347)	(0.000348)	(0.000347)	(0.000348)	(0.000338)	(0.000338)
Own Education IS							
Primary school graduates or some junior high school	0.440***	0.610***	0.616***	0.605***	0.612***	0.539***	0.543***
	(0.0816)	(0.114)	(0.115)	(0.113)	(0.114)	(0.107)	(0.108)
Junior high school graduates or some senior high school	0.696***	1.111***	1.119***	1.098***	1.106***	0.978***	0.984***
	(0.0973)	(0.124)	(0.125)	(0.124)	(0.125)	(0.118)	(0.119)
Senior high school graduates	1.333***	2.178***	2.149***	2.169***	2.139***	2.073***	2.039***
	(0.0915)	(0.113)	(0.114)	(0.113)	(0.114)	(0.108)	(0.109)
Some college or above	1.717***	2.674***	2.640***	2.654***	2.619***	2.566***	2.524***
	(0.160)	(0.164)	(0.164)	(0.164)	(0.164)	(0.162)	(0.162)
Father's Education IS							
Primary school graduates or some junior high school	-0.0200	0.0445	0.0489	0.0548	0.0593	0.105	0.109
	(0.0761)	(0.0846)	(0.0849)	(0.0846)	(0.0848)	(0.0829)	(0.0831)
Junior high school graduates or some senior high school	0.267*	0.344**	0.325*	0.330*	0.311*	0.300*	0.280*
	(0.132)	(0.131)	(0.131)	(0.131)	(0.131)	(0.130)	(0.130)
Senior high school graduates	0.00560	0.106	0.119	0.101	0.114	0.121	0.133
	(0.148)	(0.145)	(0.145)	(0.145)	(0.145)	(0.145)	(0.145)
Some college or above	0.159	0.146	0.266	0.104	0.224	0.0738	0.196
	(0.393)	(0.339)	(0.339)	(0.339)	(0.339)	(0.342)	(0.342)
Mother's Education IS							
Primary school graduates or some junior high school	0.209*	0.126	0.132	0.126	0.132	0.109	0.114
	(0.0833)	(0.0884)	(0.0885)	(0.0884)	(0.0885)	(0.0874)	(0.0874)
Junior high school graduates or some senior high school	0.0240	-0.0846	-0.0802	-0.0720	-0.0677	-0.0341	-0.0304
	(0.163)	(0.159)	(0.159)	(0.159)	(0.159)	(0.159)	(0.158)
Senior high school graduates or above	-0.298	-0.265	-0.320	-0.234	-0.289	-0.221	-0.276
	(0.181)	(0.174)	(0.174)	(0.174)	(0.174)	(0.175)	(0.174)
Urban Area	0.923***	0.608***	0.571***	0.617***	0.581***	0.631***	0.595***
	(0.0627)	(0.0718)	(0.0721)	(0.0718)	(0.0721)	(0.0702)	(0.0704)
Constant	-3.347***	-6.728***	-6.814***	-6.992***	-7.078***	-7.130***	-7.203***
	(0.553)	(0.637)	(0.639)	(0.638)	(0.640)	(0.624)	(0.626)
Province Dummy	yes	yes	yes	yes	yes	yes	yes
N	6424	6424	6424	6424	6424	6424	6424

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9: *Summary Statistics(1997): Def 0*

Variables	Informal	Formal	MeanDiff
income in last year	10.80	11.39	-0.585***
age	41.00	37.30	3.705***
education			
own	5.403	8.998	-3.596***
father's	2.154	4.057	-1.903***
mother's	1.350	2.875	-1.525***
urban	0.310	0.665	-0.355***

Table 10: *Summary Statistics(1997): Def II with 5 as firm size cutoff point*

Variables	Informal	Formal	MeanDiff
income in last year	10.85	11.47	-0.621***
age	40.35	37.48	2.864***
education			
own	5.583	9.643	-4.060***
father's	2.281	4.341	-2.061***
mother's	1.475	3.061	-1.585***
urban	0.347	0.695	-0.348***

Table 11: *Summary Statistics(1997): Def III with 5 as firm size cutoff point*

Variables	Informal	Formal	MeanDiff
income in last year	10.83	11.48	-0.647***
age	40.38	37.48	2.895***
education			
own	5.544	9.646	-4.102***
father's	2.253	4.357	-2.104***
mother's	1.452	3.076	-1.624***
urban	0.343	0.696	-0.354***

Table 12: *Summary Statistics(1997): Def IV with 5 as firm size cutoff point*

Variables	Informal	Formal	MeanDiff
income in last year	10.84	11.48	-0.643***
age	40.36	37.49	2.863***
education			
own	5.552	9.650	-4.098***
father's	2.263	4.349	-2.086***
mother's	1.461	3.066	-1.605***
urban	0.344	0.696	-0.353***

Table 13: *Summary Statistics(1997): Def V with 5 as firm size cutoff point*

Variables	Informal	Formal	MeanDiff
income in last year	10.83	11.48	-0.656***
age	40.36	37.53	2.829***
education			
own	5.517	9.659	-4.142***
father's	2.238	4.366	-2.128***
mother's	1.449	3.068	-1.619***
urban	0.341	0.697	-0.356***

Table 14: *Summary Statistics(1997): Def VI with 5 as firm size cutoff point*

Variables	Informal	Formal	MeanDiff
income in last year	10.83	11.48	-0.652***
age	40.34	37.55	2.796***
education			
own	5.525	9.663	-4.138***
father's	2.247	4.358	-2.111***
mother's	1.458	3.059	-1.601***
urban	0.342	0.697	-0.355***

Table 15: *Summary Statistics(1997): Def I with 10 as firm size cutoff point*

Variables	Informal	Formal	MeanDiff
income in last year	10.87	11.53	-0.663***
age	39.95	37.88	2.071***
education			
own	5.705	10.11	-4.402***
father's	2.386	4.477	-2.091***
mother's	1.539	3.205	-1.666***
urban	0.371	0.703	-0.332***

Table 16: *Summary Statistics(1997): Def II with 10 as firm size cutoff point*

Variables	Informal	Formal	MeanDiff
income in last year	10.87	11.53	-0.660***
age	39.93	37.90	2.031***
education			
own	5.724	10.12	-4.392***
father's	2.405	4.456	-2.051***
mother's	1.552	3.192	-1.640***
urban	0.373	0.702	-0.329***

Table 17: *Summary Statistics(1997): Def III with 10 as firm size cutoff point*

Variables	Informal	Formal	MeanDiff
income in last year	10.86	11.54	-0.673***
age	39.95	37.88	2.070***
education			
own	5.695	10.11	-4.415***
father's	2.379	4.484	-2.105***
mother's	1.534	3.208	-1.674***
urban	0.370	0.704	-0.334***

Table 18: *Summary Statistics(1997): Def IV with 10 as firm size cutoff point*

Variables	Informal	Formal	MeanDiff
income in last year	10.87	11.54	-0.671***
age	39.93	37.90	2.029***
education			
own	5.714	10.12	-4.405***
father's	2.398	4.463	-2.065***
mother's	1.548	3.195	-1.648***
urban	0.372	0.703	-0.330***

Table 19: *Summary Statistics(1997): Def V with 10 as firm size cutoff point*

Variables	Informal	Formal	MeanDiff
income in last year	10.85	11.55	-0.693***
age	39.94	37.94	1.998***
education			
own	5.653	10.12	-4.467***
father's	2.356	4.494	-2.137***
mother's	1.524	3.199	-1.676***
urban	0.366	0.705	-0.339***

Table 20: *Summary Statistics(1997): Def VI with 10 as firm size cutoff point*

Variables	Informal	Formal	MeanDiff
income in last year	10.86	11.55	-0.690***
age	39.92	37.97	1.956***
education			
own	5.672	10.13	-4.456***
father's	2.376	4.473	-2.097***
mother's	1.538	3.187	-1.649***
urban	0.368	0.704	-0.336***

Table 21: *Summary Statistics(2000): Def 0*

Variables	Informal	Formal	MeanDiff
income in last	10.79	11.18	-0.385***
age	39.91	36.32	3.586***
education			
own	6.334	9.739	-3.405***
father's	0.437	0.881	-0.444***
mother's	0.271	0.590	-0.318***
urban	0.355	0.671	-0.316***

Table 22: *Summary Statistics(2000): Def II with 5 as firm size cutoff point*

Variables	Informal	Formal	MeanDiff
income in last year	10.82	11.27	-0.451***
age	39.04	36.69	2.352***
education			
own	6.551	10.54	-3.991***
father's	0.467	0.983	-0.516***
mother's	0.297	0.654	-0.356***
urban	0.405	0.688	-0.283***

Table 23: *Summary Statistics(2000): Def III with 5 as firm size cutoff point*

Variables	Informal	Formal	MeanDiff
income in last year	10.80	11.28	-0.474***
age	39.06	36.71	2.347***
education			
own	6.498	10.56	-4.058***
father's	0.459	0.988	-0.529***
mother's	0.289	0.662	-0.372***
urban	0.401	0.690	-0.289***

Table 24: *Summary Statistics(2000): Def IV with 5 as firm size cutoff point*

Variables	Informal	Formal	MeanDiff
income in last year	10.81	11.28	-0.474***
age	39.04	36.71	2.334***
education			
own	6.522	10.55	-4.029***
father's	0.462	0.987	-0.525***
mother's	0.293	0.658	-0.365***
urban	0.403	0.689	-0.285***

Table 25: *Summary Statistics(2000): Def V with 5 as firm size cutoff point*

Variables	Informal	Formal	MeanDiff
income in last year	10.79	11.28	-0.489***
age	39.06	36.81	2.251***
education			
own	6.434	10.47	-4.039***
father's	0.451	0.975	-0.524***
mother's	0.284	0.653	-0.369***
urban	0.395	0.687	-0.292***

Table 26: *Summary Statistics(2000): Def VI with 5 as firm size cutoff point*

Variables	Informal	Formal	MeanDiff
income in last year	10.79	11.28	-0.488***
age	39.05	36.81	2.237***
education			
own	6.458	10.47	-4.008***
father's	0.455	0.974	-0.519***
mother's	0.288	0.650	-0.363***
urban	0.397	0.685	-0.288***

Table 27: *Summary Statistics(2000): Def I with 10 as firm size cutoff point*

Variables	Informal	Formal	MeanDiff
income in last year	10.83	11.33	-0.504***
age	38.74	36.97	1.770***
education			
own	6.658	11.08	-4.419***
father's	0.490	1.030	-0.540***
mother's	0.313	0.686	-0.372***
urban	0.420	0.708	-0.288***

Table 28: *Summary Statistics(2000): Def II with 10 as firm size cutoff point*

Variables	Informal	Formal	MeanDiff
income in last year	10.83	11.33	-0.498***
age	38.69	37.04	1.654***
education			
own	6.709	11.07	-4.358***
father's	0.497	1.028	-0.531***
mother's	0.319	0.682	-0.362***
urban	0.424	0.705	-0.281***

Table 29: *Summary Statistics(2000): Def III with 10 as firm size cutoff point*

Variables	Informal	Formal	MeanDiff
income in last year	10.82	11.34	-0.516***
age	38.73	36.99	1.738***
education			
own	6.649	11.07	-4.424***
father's	0.489	1.030	-0.542***
mother's	0.312	0.688	-0.376***
urban	0.419	0.709	-0.290***

Table 30: *Summary Statistics(2000): Def IV with 10 as firm size cutoff point*

Variables	Informal	Formal	MeanDiff
income in last year	10.83	11.34	-0.511***
age	38.69	37.06	1.621***
education			
own	6.700	11.06	-4.363***
father's	0.495	1.028	-0.533***
mother's	0.317	0.683	-0.366***
urban	0.423	0.706	-0.283***

Table 31: *Summary Statistics(2000): Def V with 10 as firm size cutoff point*

Variables	Informal	Formal	MeanDiff
income in last year	10.81	11.34	-0.536***
age	38.74	37.09	1.642***
education			
own	6.576	10.95	-4.374***
father's	0.479	1.016	-0.537***
mother's	0.305	0.679	-0.374***
urban	0.413	0.703	-0.290***

Table 32: *Summary Statistics(2000): Def VI with 10 as firm size cutoff point*

Variables	Informal	Formal	MeanDiff
income in last year	10.81	11.34	-0.531***
age	38.69	37.16	1.525***
education			
own	6.630	10.94	-4.308***
father's	0.486	1.014	-0.528***
mother's	0.311	0.675	-0.364***
urban	0.418	0.700	-0.283***

Table 33: *IFLS2(1997): Firm Size Cutoff is 5.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Def 0	Def I	Def II	Def III	Def IV	Def V	Def VI
ATT	0.0929*	0.173***	0.148***	0.160***	0.156***	0.174***	0.170***
	(0.0400)	(0.0357)	(0.0327)	(0.0329)	(0.0321)	(0.0342)	(0.0333)
ATU	0.218***	0.279***	0.251***	0.254***	0.250***	0.250***	0.246***
	(0.0512)	(0.0529)	(0.0512)	(0.0507)	(0.0503)	(0.0524)	(0.0520)
ATE	0.161***	0.247***	0.214***	0.220***	0.216***	0.222***	0.218***
	(0.0370)	(0.0405)	(0.0366)	(0.0356)	(0.0353)	(0.0365)	(0.0361)
OLS	0.187***	0.207***	0.204***	0.214***	0.210***	0.218***	0.215***
	(0.0307)	(0.0242)	(0.0240)	(0.0243)	(0.0242)	(0.0248)	(0.0246)
Mean	0.822***	0.891***	0.885***	0.912***	0.906***	0.923***	0.917***
Difference	(0.0603)	(0.0594)	(0.0593)	(0.0600)	(0.0599)	(0.0596)	(0.0596)
N	4270	4270	4270	4270	4270	4270	4270

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 34: *IFLS2(1997): Firm Size Cutoff is 10.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Def 0	Def I	Def II	Def IDef III	Def IV	Def V	Def VI
ATT	0.0929*	0.180***	0.167***	0.172***	0.160***	0.198***	0.176***
	(0.0400)	(0.0310)	(0.0310)	(0.0317)	(0.0316)	(0.0331)	(0.0338)
ATU	0.218***	0.246***	0.242***	0.245***	0.240***	0.249***	0.243***
	(0.0512)	(0.0581)	(0.0576)	(0.0574)	(0.0567)	(0.0558)	(0.0552)
ATE	0.161***	0.226***	0.219***	0.223***	0.216***	0.233***	0.223***
	(0.0370)	(0.0433)	(0.0431)	(0.0426)	(0.0422)	(0.0410)	(0.0408)
OLS	0.187***	0.221***	0.210***	0.220***	0.210***	0.232***	0.221***
	(0.0307)	(0.0247)	(0.0245)	(0.0249)	(0.0248)	(0.0256)	(0.0257)
Mean	0.822***	0.943***	0.933***	0.951***	0.941***	0.974***	0.964***
Difference	(0.0603)	(0.0602)	(0.0593)	(0.0606)	(0.0597)	(0.0599)	(0.0591)
N	4270	4270	4270	4270	4270	4270	4270

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 35: *IFLS3(2000): Firm Size Cutoff is 5.*

	(1) Def 0	(2) Def I	(3) Def II	(4) Def IDef III	(5) Def IV	(6) Def V	(7) Def VI
ATT	0.115***	0.194***	0.190***	0.222***	0.218***	0.228***	0.224***
	(0.0304)	(0.0315)	(0.0311)	(0.0315)	(0.0309)	(0.0325)	(0.0321)
ATU	0.181***	0.242***	0.242***	0.245***	0.242***	0.243***	0.239***
	(0.0343)	(0.0372)	(0.0372)	(0.0370)	(0.0371)	(0.0382)	(0.0383)
ATE	0.151***	0.226***	0.224***	0.237***	0.233***	0.237***	0.233***
	(0.0267)	(0.0299)	(0.0299)	(0.0300)	(0.0301)	(0.0306)	(0.0306)
OLS	0.150***	0.202***	0.201***	0.217***	0.215***	0.222***	0.220***
	(0.0244)	(0.0251)	(0.0247)	(0.0254)	(0.0250)	(0.0265)	(0.0261)
Mean	0.569***	0.688***	0.685***	0.715***	0.712***	0.729***	0.726***
Difference	(0.0477)	(0.0457)	(0.0455)	(0.0460)	(0.0458)	(0.0471)	(0.0469)
N	6424	6424	6424	6424	6424	6424	6424

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 36: *IFLS3(2000): Firm Size Cutoff is 10.*

	(1) Def 0	(2) Def I	(3) Def II	(4) Def IDef III	(5) Def IV	(6) Def V	(7) Def VI
ATT	0.115***	0.230***	0.225***	0.243***	0.237***	0.251***	0.246***
	(0.0304)	(0.0319)	(0.0309)	(0.0307)	(0.0294)	(0.0322)	(0.0305)
ATU	0.181***	0.293***	0.296***	0.292***	0.295***	0.275***	0.277***
	(0.0343)	(0.0377)	(0.0380)	(0.0377)	(0.0380)	(0.0366)	(0.0367)
ATE	0.151***	0.275***	0.276***	0.278***	0.279***	0.268***	0.268***
	(0.0267)	(0.0309)	(0.0312)	(0.0308)	(0.0311)	(0.0304)	(0.0304)
OLS	0.150***	0.242***	0.239***	0.250***	0.248***	0.258***	0.256***
	(0.0244)	(0.0245)	(0.0240)	(0.0249)	(0.0244)	(0.0263)	(0.0256)
Mean	0.569***	0.773***	0.766***	0.788***	0.781***	0.805***	0.798***
Difference	(0.0477)	(0.0442)	(0.0441)	(0.0450)	(0.0449)	(0.0458)	(0.0456)
N	6424	6424	6424	6424	6424	6424	6424

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

References

Alaniz, E., Gindling, T., and Terrell, K. (2011). The impact of minimum wages on wages, work and poverty in nicaragua. *Labour Economics*, pages 45–59.

De Soto, H. (1989). *The other path: the invisible revolution in the third world*. Number 1. New York: Harper & Row Publishers, Inc.

Dehejia, R. and Wahba, S. (1999). Causal effects in nonexperimental studies: reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94:1053 – 1062.

Dehejia, R. and Wahba, S. (2002). Propensity score matching methods for nonexperimental causal studies. *Review of Economics and Statistics*, 84(1):151 – 161.

El Badaoui, E., Strobl, E., and Walsh, F. (2010). The formal sector wage premium and firm size. *Journal of Development Economics*, 91(1):37–47.

Friedlander, B. D. and Robins, P. K. (1995). Evaluating program evaluations: new evidence on commonly used nonexperimental methods. *The American Economic Review*, 85(4):923–937.

Gong, X. and Van Soest, A. (2002). Wage differentials and mobility in the urban labour market: a panel data analysis for Mexico. *Labour Economics*, 9(4):513–529.

Ham, J. C., Li, X., and Reagan, P. B. (2011). Matching and semi-parametric IV estimation, a distance-based measure of migration, and the wages of young men. *Journal of Econometrics*, 161(2):208–227.

Ham, J. C., Li, X., and Shore-Sheppard, L. D. (2016). The Employment Dynamics of Disadvantaged Women: Evidence from the SIPP. *Journal of Labour Economics*, 34(4):899 – 944.

Harris, J. R. and Todaro, M. P. (1970). Migration, unemployment and development: a two-sector analysis. *The American Economic Review*, pages 126–142.

Heckman, J. J., Ichimura, H., Smith, J., and Todd, P. E. (1998). Characterizing selection bias using experimental data. *Econometrica*, 66(5):1017 – 1098.

Heckman, J. J., Ichimura, H., and Todd, P. E. (1997). Matching as an econometric evalutaion estimator: evidence from evaluating a job training programme. *Review of Economic Studies*, 64(4):605–654.

Hohberg, M. and Lay, J. (2015). The impact of minimum wages on informal and formal labor market outcomes- evidence from Indonesia.

La Porta, R. and Shleifer, A. (2008). The unofficial economy and economic development. *Brookings Papers on Economic Activity*, pages 275–236.

Lewis, W. A. (1954). Economic development with unlimited supplies of labour. *Manchester School of Economic and Social Studies*, 22(2):139–191.

Maloney, W. F. (2004). Informality revisited. *World Development*, 32(7):1159–1178.

Michalopoulos, C., Bloom, H., and Hill, C. (2004). Can propensity score methods matchs the findings from a random assignment evaluation of mandatory welfare-to-work programs? *The Review of Economics and Statistics*, 86(1):156 – 179.

Nelson, E. G. and Bruijn, E. J. D. E. (2005). The voluntary formalization of enterprises in a developing economy: the case of Tanzania. *Journal of International Development*, 593(4):575–593.

Pratap, S. and Quintin, E. (2006). Are labour markets segmented in developing countries? A semiparametric approach. *European Economic Review*, 50(7):1817–1841.

Rosenbaum, P. and Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Smith, J. A. and Todd, P. E. (2005). *Does matching overcome LaLonde's critique of nonexperimental estimators?*, volume 125.

Staneva, A. and Arabsheibani, G. R. (2014). Is there an informal employment wage premium? Evidence from Tajikistan. *IZA Journal of Labour & Development*, 3(1):1.

Strauss, J., Witoelar, F., and Sikoki, B. (2016). The fifth wave of the Indonesia Family Life Survey : overview and field report.