

Baskaya, Yusuf Soner; Hulagu, Timur

**Conference Paper**

## Informal-Formal Worker Wage Gap in Turkey: Evidence From A Semi-Parametric Approach

51st Congress of the European Regional Science Association: "New Challenges for European Regions and Urban Areas in a Globalised World", 30 August - 3 September 2011, Barcelona, Spain

**Provided in Cooperation with:**

European Regional Science Association (ERSA)

*Suggested Citation:* Baskaya, Yusuf Soner; Hulagu, Timur (2011) : Informal-Formal Worker Wage Gap in Turkey: Evidence From A Semi-Parametric Approach, 51st Congress of the European Regional Science Association: "New Challenges for European Regions and Urban Areas in a Globalised World", 30 August - 3 September 2011, Barcelona, Spain, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/120000>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Informal-Formal Worker Wage Gap in Turkey: Evidence From A Semi-Parametric Approach

Yusuf Soner Başkaya      Timur Hulagu<sup>1</sup>

This version: February 25, 2011.

(PRELIMINARY AND INCOMPLETE)

## Abstract

Using individual level data from Turkstat Household Labor Force Survey for 2005-2009 period and a variety of parametric and semi-parametric techniques, we test two hypothesis regarding formal and informal labor markets: whether there is a wage gap between formal and informal workers and whether this gap is sensitive to variations in unemployment rates across regions and over time, where the formality of employment is defined with respect to registry status of the individuals to compulsory Social Security System. In line with most studies, the formal workers earn more than informal workers, as suggested by standard wage regressions, conditional on workers' observed individual characteristics. On the other hand, considering the limitations of parametric methods and possibility of misleading results due to the different distributional characteristics of formal and informal workers, we alternatively implement propensity score matching. In contrast with the recent studies for other developing countries showing that the wage gap estimates with propensity score matching are insignificant, we do find a large and sizeable wage gap between formal and informal workers in Turkey. While parametric methods give similar estimates for formal-informal wage gap within gender groups, the semi-parametric estimates suggest that the observed formal-informal wage gap is larger among females compared to males. Finally, we show that although the parametric methods suggest that formal wage premium increases with higher unemployment rates, the semi-parametric methods show that these gaps are insensitive to unemployment rate variations across regions or over time.

Keywords: Formal/Informal Employment, Wage Gap, Propensity Score Matching, Regional Labor Markets.

JEL classification: C14; J30; J42; J60; O17

---

<sup>1</sup>Research and Monetary Policy Department, Central Bank of Turkey, Ankara, 06100, Turkey. e-mail: soner.baskaya@tcmb.gov.tr, timur.hulagu@tcmb.gov.tr. The views expressed in this paper are those of authors and do not necessarily reflect the official views of the Central Bank of Turkey.

# 1 Introduction

One of the distinctive characteristics of labor markets in developing countries is the mass number of workers working under informal employment contracts. As stated by Freeman (2007), the persistence of large informal sectors in developing countries puts a premium on increasing our knowledge of how informal sector labor markets work and finding institutions and policies to deliver social benefits to workers in that sector. According to OECD (2009), informal employment constitutes to an average of 57% of total employment across countries in Latin America while that rises to 70% for countries in South and Southeast Asia.<sup>2</sup>

A particular hypothesis regarding the coexistence of formal and informal employment is whether the labor markets are characterized by a dualistic structure, such that workers in informal jobs are those who do not have access to jobs in the formal sector. Traditional dual labor market theories, starting with Lewis (1954), describe the informal sector as the disadvantaged sector. In particular, the workers who enter the informal market are those who are rationed out of the formal sector due to wages above market-clearing prices (Harris and Todaro, 1970; Stiglitz, 1976). In other words, the workers with no access to formal jobs may have limited options, as a result of which they accept jobs with lower wages, worse working conditions and/or without access to social security coverage. In other words, the entry barriers to the formal jobs may lead to a wage gap between formal and informal workers who have comparable characteristics.

The existence of formal wage premium has been documented by various studies relying on parametric techniques (see, for example, Mazumdar, 1981; Heckman and Hotz, 1986; Roberts, 1989; Pradhan and Van Soest, 1995; Tansel, 1999; Gong and Van Soest, 2001). However, due to the possibility that the distribution of observable characteristics of formal and informal workers differ from each other, the recent literature questioned the suitability of the parametric methods for estimating wage gaps with parametric methods and challenges the results obtained with them. For example, Pratap and Quintin (2006) shows that the formal-informal wage gap observed in Argentina, which is found to be significant and large with parametric methods, disappear once the semi-parametric methods such as propensity score matching is used. This result can also be regarded as a suggestive evidence that the earlier findings in favor of dual labor markets in developing countries may be an artifact of utilization of inappropriate estimation techniques.

Using a variety of parametric and semi-parametric techniques and individual level data from Turkstat Household Labor Force Survey (THLFS) for 2005-2009 period, this study tests whether there is a wage gap between formal and informal in Turkey. The Turkstat Household Labor Force Survey (THLFS) provides information on large set of demographic

---

<sup>2</sup>Source: OECD Development Center “Is Informal Normal? Towards More and Better Jobs in Developing Countries” (2009), based on ILO LABORSTA database and ILO Global Employment Trends Report, 2009.

and job-related individual characteristics, including whether the individual is registered to the legally mandatory *Sosyal Guvenlik Kurumu* (i.e. Turkish Social Security Institution, SGK hereafter) at his current job. The registration status to SGK allows us to determine the formality status of individuals' employment, in line with Maloney (2004) and Ramos et al (2010). In particular, we consider workers formally employed if they are registered to compulsory SGK in Turkey.

Our main results are as follows: In line with most studies, standard Mincerian wage regressions suggest that the formal workers earn more than informal workers in Turkey, conditional on workers' observed individual characteristics. On the other hand, considering the limitations of parametric methods that would raise possibility of misleading results, we also estimate the formal-informal wage gap by using propensity score matching (PSM), which is a semi-parametric method that requires no assumption on the form of earning functions. PSM also considers the possibility that the individual characteristics of formal and informal workers may have different distributional characteristics. We find a large and sizeable wage gap between formal and informal workers in Turkey also with PSM. It is worth noting that these results are robust to different matching techniques. However, it is important to note that while our results overcome potential problems associated with parametric estimation, they may also be subject to some potential shortcomings due to being unable to fully control for the potential effect of unobserved skills between formal and informal workers in our PSM exercise where we match on the observed characteristics.

Our results provide empirical support for the existence of segmented labor markets for formal and informal workers in Turkey, both with parametric and semi-parametric techniques. While this is in line with findings for the developing countries based on the parametric techniques, our analysis provides different result from the studies which argue that the semi parametric estimation methods overturns the empirical support for the segmented labor markets.

We also estimate wage gap between formal and informal workers with respect to gender types. While standard regression results indicate that males and females have similar formal wage premium, PSM gives smaller estimates for the formal wage premium for the males and larger estimates for the females in Turkey. Moreover, PSM suggests that there is a significant gap across males and females in terms of the magnitude of the formal-informal wage gap.

Besides testing whether there is a significant wage gap between observationally comparable formal and informal workers, which would provide evidence on existence of segmented labor markets for these worker types, we also analyze whether the wage gaps between formal and informal workers are sensitive to the variations in the labor market activity across years in our sample period and 26 NUTS regions. While a relatively small number of studies have previously looked at whether there is a difference between formal and informal workers in

terms of sensitivity of wages to unemployment variations, such an analysis deserves attention for various reasons. On the public policy side, the answer for this question has important implications for whether poverty gap between the formal and informal workers widen during economic downturns. In a related manner, this has important implications for the public policies aiming at providing social benefits to the sectors with different employment types.

First, we estimate separately wage curves for formal and informal workers and document that there are significant differences across these groups in the sensitivity of wages to variations in regional unemployment rates.<sup>3</sup> Finally, as an alternative exercise, we first construct a panel data for formal-informal wage gaps from estimated gaps for each year and NUTS region with PSM, and see test the sensitivity of this measure to variations in unemployment rate across regions and over time. This specification shows that estimated wage gaps are sensitive to unemployment rates when we do not control for the year fixed effects, but insensitive to the specification with time fixed effects, indicating that mixed evidence on whether the formal-informal wage gap varies with the fluctuations in the labor market.

The rest of the paper is structured as follows. Section 2 briefly introduces the main features of our dataset. In section 3, we discuss our empirical strategy to estimate the informal wage gaps, their sensitivity to variations in labor market activity across regions and time as well as our empirical estimates. Section 4 concludes the paper.

## 2 Data

The data set used in this study is taken from annual individual data releases of the Turkstat Household Labor Force Survey (THLFS) for the 2005-2009 period.<sup>4</sup> As we are interested in how hourly wages of individuals respond to aggregate variations in the unemployment rates, we exclude unpaid family workers, self-employed individuals and individuals stated as employers in the survey from the sample. Also, due to possible important measurement problems about their earnings, individuals younger than 15 years of age are excluded from the sample. Finally, following OECD(2009), we focus on the workers employment in non-agriculture sector, as it is hard to distinguish between formal and informal employment agriculture sector. Therefore, we mostly focused on the sample which excludes workers in agriculture sector<sup>5</sup>. In all of our regressions, we use the population weights provided by

---

<sup>3</sup>In a recent paper, Ramos et al. (2010) found that there exists a high negatively sloped wage curve in Colombia for informal workers, but not for formal workers.

<sup>4</sup>All private households who are living in the territory of the Republic of Turkey are covered by this annual survey. Residents of schools, dormitories, kindergartens, rest homes for elderly persons, special hospitals, military barracks and recreation quarters for officers are not covered by this survey. For more information, see the Turkstat website.

<sup>5</sup>See Appendix Table 1 for the sample exclusion rules.

Turkstat.

Although the THLFS provides the individual level data on a wide range of the individuals' demographic and job-related characteristics starting from 2002, we need to focus on the post-2005 period, as it is impossible to compute the hourly wages of individuals due to the absence of data on usual hours worked by the individuals in 2002-2004 period.

The data on real hourly wages is obtained by dividing the monthly nominal after tax cash earnings by the total hours worked in the month. It is then deflated by regional prices, provided by Turkstat into 2008 prices. For measuring the variations in the aggregate labor market activity, we use the regional unemployment rates,  $U_{rt}$  provided by Turkstat. Due to measurement problems for agricultural workers, we use non-agricultural unemployment rates and present results pertaining to exclusion of agricultural workers.<sup>6</sup>

### 3 Empirical Strategy and Results

#### 3.1 Parametric tests

In this section, we estimate a Mincerian wage regression, which takes into account the possibility that the determinants of the hourly wages differ across the formal and informal workers:

$$\log W_{irt} = \alpha + \beta F_{irt} + X'_{irt}\gamma + \mu_r + \lambda_t + \nu_{irt} \quad (1)$$

where  $W_{irt}$  is the real hourly wage rate of worker  $i$  observed in region  $r$  at time  $t$ .  $F_{irt}$  is the formality status of the worker as explained below in detail.  $F_{irt}$  takes value 1 if the worker is categorized as formal.  $X_{irt}$  represents the set of measured characteristics of worker  $i$ ,  $\mu_r$  is a region effect,  $\lambda_t$  is a time effect and  $\nu_{irt}$  is the error term. Other variables which are used to control for individual heterogeneity are age, gender, marital status, employment location, years of education, enrollment to a school, years of tenure at the firm, firm size, industry of the firm according to the NACE Rev.1 classification, occupational group according to the ISCO-88 classification, permanency of the job, part-time work, other activity to earn income and employment status in the same month of last year. These variables are explained further in detail in the appendix section.

One of our variables of interest is the informality status of the individuals' employment. The informal employment can be divided into a number of subcategories, i.e. informality might arise from the nature of the institution that the individual work for (informal sector enterprises), from the jobs in the formal sector which are unprotected, or from the households producing goods exclusively for their own final use and households employing

---

<sup>6</sup>However, our results are robust to the inclusion of agricultural workers and agricultural unemployment rates in the sample.

paid domestic workers. Regarding this difficulty and seeking for a standardization across countries, the International Labour Organization (ILO) defines informal sector employment as self-employed; wage workers in insecure and unprotected jobs (unregistered, casual, temporary); and household workers (see Freeman (2007) and ILO).<sup>7</sup> However, some of the subcategories in ILO's are not associated with wage-earning for individuals, making such a categorization unsuitable for the analysis of how informality affects the wages. Therefore, we use the alternative definition by OECD (2009), which defines informal employment using whether the individual is covered by the social security system or not. In particular, we categorize a worker as an informal worker if the worker is not registered to Social Security Institution at his current job.<sup>8</sup>

The main parameter of interest in Equation (1) is  $\beta$ , where  $\beta > 0$  implies that formal workers earn higher than informal workers, after controlling for observed individual characteristics. Pratap and Quintin (2006), for example, report the estimates for  $\beta$  for Argentina in the 0.23-0.37 range. Using Turkstat's 1994 Household Expenditure Survey, Tansel (1999) finds that formal male workers earn 68 percent higher than informal male workers and formal females earn 2.5 times more the informal females.

Following the large body of existing literature, we first provide results obtained with standard regression analysis. In particular, we estimate Equation (1) using individual level data from Turkish Household Labor Force Survey for 2005-2009 period. Table 3 presents the estimation results for three different definitions of our sample.

The column 1 presents the results with all individuals including both agriculture sector wage workers and the workers in the community services sector, which mostly correspond to civil servants. First, it is worth noting that the parameter estimates for the wage returns to individual characteristics, such as age, education, tenure, marital status etc., are in line with the expectations based on earlier theoretical and empirical findings on determinants on wages. The main parameter of interest, i.e. the wage difference between formal and informal workers, is estimated as 20 percent on average, which is significant at 1% level. When we exclude the agricultural workers from the sample, due to the possible measurement problems of the employment characteristics of the workers in agriculture sector workers, we still find a wage gap around 19.3 percent. Finally, we estimate the formal-informal worker wage gap after excluding workers in community services sector and agriculture sector. This leaves us with a sample that can be regarded as non-agricultural sector workers in the

---

<sup>7</sup>On the other hand, the informality criteria are not adapted for agriculture sector and it is hard to distinguish formal and informal agriculture. Therefore, we mostly focused on the sample which excludes workers in agriculture sector.

<sup>8</sup>This approach is similar to empirical studies by Maloney (2004) and Ramos et al. (2010). Nevertheless, our results are robust to the choice of the definition of informality (i.e. we obtain the same conclusions once we add casual or temporary workers to the definition of informal workers. Currently, we just control for casual and temporary workers).

private sector. With this sample, we still find an hourly wage gap of 15.9 percent between formal and informal workers in Turkey.

### 3.2 Semiparametric tests

Parametric methods may suffer from a possible misspecification of earning functions especially when distributions of observed individual characteristics are different for formal and informal workers. Propensity Score Matching (PSM), on the other hand, can overcome this problem by assessing wage gaps for similar workers, matched with each other using information on observed characteristics. This two-step semiparametric approach does not make any assumption on the earning function. In particular, a probit estimation of propensity scores is applied in the first step. Given a set of observed characteristics, propensity score measures the probability that the individual works in the formal sector. In the second step, wages of workers with the similar propensity score are matched. There are several methods to determine the “similarity” of the propensity scores<sup>9</sup> and we pursue the most used two, namely caliper matching and nearest neighbor matching.

Formally, following LaLonde (1986) and Heckman et al. (1999), we estimate the formal wage gap as the average effect of treatment on the treated (ATT):

$$\beta = E(w^F|X, Formality = 1) - E(w^I|X, Formality = 1) \quad (2)$$

where  $X$  is the observable characteristics as defined above while  $w^F$  and  $w^I$  are the formal and informal real hourly wage rates, respectively. By definition, the second term is unobservable. However, as suggested by Rosenbaum and Rubin (1983, 1984), if the formality selection only occurs with respect to characteristics  $X$  then one can use the matching estimator  $\beta^m$  as

$$\beta^m = \frac{1}{N} \sum_{i \in F} \left( w_i^F - \sum_{j \in I} \eta_{ij} w_j^I \right) \quad (3)$$

where  $\eta_{ij}$  is the weight of informal worker  $j$  for comparison with formal worker  $i$  and  $N$  is the number of formal workers in the sample. Weights  $\eta_{ij}$  are determined by several methods and any method should have the property that as the difference between propensity scores increase the weight falls. The first method we use is caliper matching, where an individual  $i$  with the propensity score  $p_i$  is only matched with an individual  $j$  with the propensity score  $p_j$  if  $|p_i - p_j| < \delta$ . The maximum distance  $\delta$  is chosen to be  $10^{-4}$  as in Pratap and Quintin (2006), who also uses similar techniques to estimate wage gap between formal and informal

---

<sup>9</sup>See Caliendo and Kopeinig (2005) for a comprehensive discussion on the implementation of PSM and various matching algorithms.



workers in Argentina. However, we obtain similar qualitative results with  $\delta = 10^{-3}$ . More formally, in caliper matching we use the following weights:

$$\eta_{ij} = \begin{cases} 0 & \text{if } |p_i - p_j| > \delta \\ \frac{\frac{1}{|p_i - p_j|}}{\sum_{\{i, j: |p_i - p_j| \leq \delta\}} \frac{1}{|p_i - p_j|}} & \text{otherwise.} \end{cases} \quad (4)$$

As an alternative to caliper matching, we check the robustness of our results to the matching method by using nearest neighbor matching, where wage of each formal worker is compared to that of informal worker with the  $n$  closest propensity score. We follow Pratap and Quintin (2006) also for the choice of  $n$ , where we take  $n = 1$  for the baseline. However, we obtain similar results also with different choices of  $n$ , such as 2 or 5.

As stated above, the major concern for estimating the wage gaps between formal and informal workers is the possible misspecification problem due to disregarding the possible differences across the earnings function of the workers in these categories, as well as the possibility of differences in the observed characteristics of the formal and informal workers. This point has recently been shown to have important implications for the estimated wage gap between formal and informal workers in Argentina by Pratap and Quintin (2006). In particular, Pratap and Quintin (2006) challenges the estimates large wage gaps between formal and informal workers obtained with standard regressions, and show that non-parametric methods, such as propensity score matching, yield insignificant wage gaps for Argentina. They interpret these results as indicative for the role of misspecification in earlier literature in providing empirical support for the hypothesis that the labor markets for the formal and informal workers in the developing countries are segmented.

Tables 5 and 6 provide results from two different methodologies mentioned above, i.e. caliper matching and nearest neighborhood matching respectively. However, it may be first useful to look at determinants of formality status of the individuals. Table 4 presents the probit marginal effects for determinants of the formality status by each year in our sample. It is important to note that the determinants of the formality status are mostly in line with our expectations. For instance, we find that the probability of being a formal worker increases monotonically with more years of education. In addition, while the results differ slightly by years, the males have at least 20 percent more likely to be formal workers than the female. We also find that the married individuals are less likely to be informal workers. The probability of being a formal worker increases, though at a decreasing rate, with age and tenure status. Finally, we find that the probability of informal employment decreases with the size of the firm that the individual is employed at and that there is a heterogeneity across occupation groups in terms of the formality status.

The results presented in Tables 5 and 6 show the wage gap between formal and informal workers by years and NUTS regions. The estimation of the wage gaps without regional decomposition is presented in the last lines of Tables 5 and 6, which present results from two different matching methods. These results suggest that although the wage gap changed over time, it has been significant for all years. For example, while the formal-informal wage gap has been more than 20 percent for 2005 and 2009, they have been around 10-15 percent in the remaining years. Finally, although the regional decompositions by years and NUTS regions indicate wage gaps at different magnitudes, they mostly indicate significant wage gaps observed at the regional level as well.

### **3.3 Formal-Informal Wage Gaps By Gender Groups**

We also analyze the magnitude of formal-informal worker wage gap by the gender groups. The upper panel of Table 7 presents the results from standard regression techniques obtained separately with the male and female samples. These indicate that the formal-informal wage gaps within males and females are 18.8 percent and 19.9 percent respectively, which are significant at 1 percent level. These indicate that the wage premium for being a formal worker do not significantly vary by gender types. On the other hand, with semiparametric techniques, we find that the wage gap for males is around 15 percent, whereas the wage gap for the females is 25 percent. Moreover, we find that these formal-informal wage gap for the females is statistically higher than that for the males. These suggest that parametric techniques reveal misleading results in terms of the magnitude of the wage premium due to formal employment for females relative to males.

### **3.4 Formal-Informal Wage Gaps and Variations in Labor Market Activity By Regions and Time**

A particular observation for the formal-informal worker wage gaps in Tables 5 and 6 is that they vary by regions and years. Based on this observation, we ask whether the variations in labor market activity across time and regions explain the changes in the wage gap. For example, if the workers in the informal jobs have relatively less bargaining power during the periods of low labor market activity, we may observed higher wage gaps between formal and informal workers.

The analysis of the sensitivity of formal-informal worker wage gaps deserves attention for various reasons. On the public policy side, the answer for this question has important implications for whether poverty gap between the formal and informal workers widen during economic downturns. In a related manner, this has important implications for the public policies aiming at providing social benefits to the sectors with different employment types. Such an analysis may provide important guidelines for the design of comprehensive labor

market reforms aiming at reducing the size of informal labor markets while increasing the flexibility of the labor markets. For example, a case where the main source of wage flexibility is the informal employment may suggest that policies decreasing the extent of the informal employment may be accompanied with the policies inducing more flexible labor markets in formal and/or informal labor markets.

One way of testing whether such wage gaps vary with the overall labor market activity, is to estimate wage curves for the formal and informal workers separately.<sup>10</sup> The results in Table 8 show that while the wages of informal workers display a significant decline with higher unemployment rates, the wages of the formal workers do not show a significant variation over the variations in the labor markets. In particular, we find that the unemployment elasticity of hourly wages of informal workers is -0.24 , whereas the estimates for the formal workers is -0.04 and insignificant, suggesting that the formal-informal wage gap widens during the periods of low labor market activity. These results also suggest that the wage curve observed for Turkey, which has recently been documented by Baltagi et al. (2011), exists mainly for the informal workers.

On the other hand, as stated above, the possibility that the formal and informal workers may differ from each other in terms of the observed and unobserved characteristics, the differences among them in terms of the sensitivity of their wages to the unemployment variations may not be attributed to the formality status. As an alternative to standard wage-curve estimation, we look at how average regional formal-informal wage gaps by years obtained from matched individuals vary with the changes in the unemployment rates. This exercise potentially reveals more information about the role of formality status in explaining the changes in the wage gaps with the unemployment fluctuations. In particular, we estimate:

$$\text{wage\_gap}_{rt} = \beta_0 + \beta_1 \log U_{rt} + \lambda_r + \mu_t + e_{rt} \quad (5)$$

where  $\text{wage\_gap}_{rt}$  is the wage gap between informal and formal workers estimated with caliper matching for region  $r$  and time  $t$ ,  $\log U_{rt}$  is the natural logarithm of nonagricultural unemployment rate,  $\lambda_r$  is the region fixed effects and  $\mu_t$  is the year fixed effects.

The results presented in Table 9 show that the empirical support for the sensitivity of wage gaps to the unemployment rates is sensitive to the controls for the region fixed effects, which can be regarded as unobserved region specific factors that might have affected the relative earnings of informal workers. In summary, our exercise using the formal-informal wage gaps from matched individuals for 2005-2009 period and 26 NUTS regions do not

---

<sup>10</sup>See Blanchflower and Oswald (1990, 1994) for the idea of wage curves. Recently Baltagi et al. (2011) show the existence of wage curve for Turkey, where the unemployment elasticity of hourly wages has been estimated to be -0.1. Finally, Ramos et al. (2010) uses wage curve methodology to estimate whether the wage of formal and informal workers have different sensitivity to variations in regional unemployment rates.

provide support for sensitivity of wage gaps to unemployment variations.

## 4 Conclusion

In this paper, we use a rich individual level data from Turkstat Household Labor Force Survey (THLFS) for 2005-2009 period and a variety of parametric and semi-parametric techniques to answer two different questions regarding formal and informal labor markets: whether there are wage gaps between formal and informal workers and whether these gaps are sensitive to variations in unemployment rates across regions and over time. Our analysis on the basis of standard Mincerian regressions indicate that the formal workers earn more than informal workers conditional on individuals' observed individual characteristics. When we use propensity score matching by considering the possibility of misleading results due to the different distributional characteristics of formal and informal workers, we still find significant wage gaps with magnitudes comparable to the regression results. This contrasts with the recent studies for other developing countries, which find no wage gaps with semi-parametric techniques, suggesting that empirical evidence for significant wage gaps between formal and informal workers is an artifact of parametric techniques. These results can be regarded as a support for the existence of dual labor markets in Turkey among formal and informal workers, where individuals with similar observational characteristics face different wages under formal and informal employment contracts in Turkey.

On the other hand, we show that parametric techniques provide misleading results for magnitude of the formal-informal wage gaps by gender types and the sensitivity of the formal-informal wage gaps to the variations in the labor market activity across regions in Turkey. We find that the returns to becoming a formal worker for females is almost twice as much as that for the males with PSM, while parametric estimates indicate similar returns. Finally, although the parametric methods suggest that formal wage premium increases with higher unemployment rates, the semi-parametric methods show that these gaps are insensitive to unemployment rate variations across regions or over time.

## References

- [1] Baltagi, B.H., Baskaya, Y.S., Hulagu, T., 2011. The Turkish wage curve: Evidence from the household labor force survey. Unpublished manuscript.
- [2] Blanchflower, D.G., Oswald, A.J., 1990. The wage curve. *Scandinavian Journal of Economics* 92, 215-235.
- [3] Blanchflower, D.G., Oswald, A.J., 1994. *The Wage Curve*. MIT Press, Cambridge MA.

- [4] Caliendo, M., Kopeinig, S., 2005. Some practical guidance for the implementation of propensity score matching. Paper 1588. Institute for the Study of Labor, IZA.
- [5] Freeman, R.B., 2007. Labor regulations, unions, and social protection in developing countries: Market distortions or efficient institutions?. *Handbook of Development Economics* 5, 4657-4702, Edited by: Dani Rodrik and Mark Rosenzweig.
- [6] Gong, X., Van Soest, A., 2001. Wage differentials and mobility in the urban labor market: A panel data analysis for Mexico. IZA, Bonn Discussion Paper No. 329.
- [7] Harris, J.R. and Todaro, M.P., 1970. Migration, unemployment and development: A two-sector analysis. *American Economic Review*, Vol. 60(1), pp.126-142.
- [8] Heckman, J.J., Hotz, V., 1986. An investigation of labor market earnings of Panamanian Males. *Journal of Human Resources* 21, 507-542.
- [9] Heckman, J.J., Lalonde, R.J., Smith, J., 1999. The economics and econometrics of active labor market programs. In: Ashenfelter, O., Card, D. (Eds.), *Handbook of Labor Economics*, vol. 3.
- [10] ILO (International Labor Organization), 2009. *Global Employment Trends Report, 2009*. Geneva: ILO.
- [11] LaLonde, R.J., 1986. Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review* 76, 604-620.
- [12] Lewis, W.A., 1954. *Economic Development with Unlimited Supplies of Labour*. Manchester School, Vol. 22(2), pp.139-91.
- [13] Maloney, W.F., 2004. Informality revisited. *World Development* 32, 1159-1178.
- [14] Mazumdar, D., 1976. The rural-urban wage gap, migration, and the shadow wage. *Oxford Economic Papers* 28, 406-425.
- [15] Mazumdar, D., 1976. The urban informal sector. *World Development* 4, 655-679.
- [16] Mazumdar, D., 1981. *The urban labor market income distribution: A study of Malaysia*. Oxford University Press, Oxford.
- [17] Mincer, J., 1974. *Schooling, Experience and Earnings*. Columbia University Press, New York. *The Review of Economics and Statistics* 72, 334-338.
- [18] OECD, 2009. *Is Informal Normal? Towards More and Better Jobs in Developing Countries*. Jutting, J., de Laiglesia J.R. (eds.), OECD Development Centre Studies, Paris.

- [19] Pradhan, M., Van Soest, A., 1995. Formal and informal sector employment in urban areas of Bolivia. *Labor Economics* 2, 275-297.
- [20] Pratap, S., Quintin, E., 2006. Are labor markets segmented in developing countries? A semiparametric approach. *European Economic Review* 50, 1817-1841.
- [21] Ramos, R., Duque, J.C., Surinach, J., 2010. Is the wage curve formal or informal? Evidence for Colombia. *Economics Letters* 109, 63-65.
- [22] Roberts, B.R., 1989. Employment structure life cycle and life chances: Formal and informal sectors in Guadalajara. In: Portes, A., Castells, M., Benton, L.A. (Eds.), *The Informal Economy: Studies in Advanced and Less Developed Countries*. Johns Hopkins University Press, Baltimore.
- [23] Rosenbaum, P., Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, 41-55.
- [24] Rosenbaum, P., Rubin, D.B., 1984. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association* 79, 516-524.
- [25] Stiglitz, J.E., 1976. The Efficiency Wage Hypothesis, surplus Labor, and the Distribution of Labour in LDCs. *Oxford Economic Papers*, Vol. 28(2), pp.185-207.
- [26] Tansel, A., 1999. Formal versus informal sector choice of wage earners and their wages in Turkey. *Economic Research Forum Working Paper No. 9927*.
- [27] Türkiye İstatistik Kurumu (Turkish Statistical Institute, Turkstat). Household Labor Force Survey Data 2005-2009.

## A Data Appendix

In this appendix, we provide details about our dataset. First, we present our data coverage and number of observations for different restrictions in Table 1.

Second, we summarize our data with respect to informality for different subgroups. Particularly, Table 2 lists percentages of formal and informal workers for four individual characteristic categorizations and three different sample coverage.

Finally, we give details about individual specific control variables that we use. Following Mincer (1974), we regress our dependent variable on a number of control variables related to individual heterogeneity, which are listed below:

- Age. The survey provides eleven age categories in 5-year intervals.

Table 1: Number of Observations

Restriction/Selection Rule	Observations
All observations in sample years 2005 to 2009	2,453,265
Civilian wage workers age 15 and over, with positive sampling weight, formality status and non-missing demographics such as: age, tenure, gender, marital status, education etc.	383,280
Excluding:	
Individuals with no wage information	379,512
Individuals in agricultural sector	367,095
Main Sample (Excluding individuals in agricultural sector):	
Male	286,034
Female	81,061
Alternative Sample (Including individuals in agricultural sector):	
Male	294,169
Female	85,343

- Gender. Female=1 and Male=0.
- Marital status. Two dummy variables are constructed for marital status. First, Single=1 for individuals who never been married, and zero otherwise. Second, Married=1 for individuals who are currently married and living together, and zero otherwise.
- Employment location. Urban=1 and Rural=0.
- Education. The variable *educ* is years of completed education, while the variable *enrolled* is a binary variable which takes the value 1 for individuals enrolled to a school, and zero otherwise. Variable *req\_att* equals to 1 for individuals who are enrolled in a school that requires regular attendance, 0 otherwise.
- Social security registration: Binary variable which takes the value 1 if the individual is registered in the social security administration, and zero otherwise.
- The individual's years of tenure at the firm. This is calculated as the starting year at the current job subtracted from the survey year.
- Industry classification. This is a set of 9 binary variables categorized according to the NACE Rev.1 classification pertaining to the industry. They include agriculture, mining, manufacturing, electricity, construction, transportation, trade and finance, and community, social and personal services.
- Occupational group. This is a set of 9 binary variables categorized according to the ISCO-88 classification. They include legislators, senior officials and managers; professionals; technicians and associate professionals; clerks; service workers and shop and

Table 2: The Fraction of Formal Workers By Types

	All workers		All but agricultural workers		All but community and agricultural workers	
	Formal	Informal	Formal	Informal	Formal	Informal
Gender						
Male	72.89%	27.11%	74.44%	25.56%	69.12%	30.88%
Female	72.39%	27.61%	75.97%	24.03%	71.89%	28.11%
Age						
Old	75.5%	24.5%	77.94%	22.06%	71.03%	28.97%
Young	70.46%	29.54%	72.12%	27.88%	68.69%	31.31%
Tenure						
High	85.04%	14.96%	87.06%	12.94%	80.17%	19.83%
Low	65.48%	34.52%	67.43%	32.57%	65.46%	34.54%
Education						
High	88.58%	11.42%	88.82%	11.18%	83.43%	16.57%
Low	58.57%	41.43%	61.45%	38.55%	60.62%	39.38%

Notes:

Young (old) refers to individuals younger (older) than sample mean value for years of age, which is 34.1. Low (high) tenure refers to individuals with tenure less (more) than the sample mean value, which is 6.94 years. Low (high) education refers to individuals with less than or equal to 8 years of schooling (more than 8 years of schooling).

market sales workers; skilled agricultural and fishery workers; craft and related trades workers; plant and machine operators and assemblers; and elementary occupations.

- Permanency of the job. Permanent=1, and Temporary or Seasonal=0.
- Employment type. Full-time=0 and part-time=1.
- Other activity to earn income. Yes=1 and no=0.
- Firm size. This is measured by the number of persons employed in the firm and summarized by 5 binary variables corresponding to the following categories: less than 10 employees, 10-24, 25-49, 50-249, 250-499, and 500 and more.
- Employment status in the same month of last year. Binary variable which takes the value 1 if the individual was working in the same month of last year, and zero otherwise.



Table 3: The Formal/Informal Wage Gap Estimated with Mincerian Wage Regression

	All workers	All but agricultural workers	All but community and agricultural workers
Formality	0.201 (0.038)***	0.193 (0.036)***	0.159 (0.028)***
Age	0.040 (0.002)***	0.042 (0.002)***	0.036 (0.001)***
Age <sup>2</sup>	-0.0005 (0.00002)***	-0.0005 (0.00002)***	-0.0004 (0.00002)***
Gender	-0.101 (0.011)***	-0.096 (0.011)***	-0.104 (0.017)***
Marital Status			
Single	-0.070 (0.004)***	-0.072 (0.005)***	-0.058 (0.009)***
Married	0.018 (0.006)**	0.016 (0.006)*	0.020 (0.007)**
Req_att	-0.079 (0.029)**	-0.079 (0.029)**	-0.072 (0.032)*
Urban	0.021 (0.010)*	0.018 (0.010)	0.028 (0.009)**
Enrolled	0.057 (0.012)***	0.060 (0.013)***	0.015 (0.011)
Permanent	0.078 (0.019)***	0.070 (0.019)***	0.054 (0.020)**
Part-time	0.407 (0.025)***	0.408 (0.026)***	0.321 (0.047)***
More than 1 job	-0.044 (0.008)***	-0.050 (0.009)***	0.044 (0.013)***
Employed last year	0.079 (0.010)***	0.081 (0.010)***	0.076 (0.012)***
Experience	0.019 (0.001)***	0.019 (0.001)***	0.016 (0.002)***
Experience <sup>2</sup>	-0.0002 (0.00004)***	-0.0002 (0.00004)***	-0.0001 (0.00007)
Firm size			
10 to 24	0.164 (0.011)***	0.172 (0.013)***	0.123 (0.009)***
25 to 49	0.179 (0.015)***	0.182 (0.016)***	0.144 (0.012)***
50 to 249	0.248 (0.020)***	0.250 (0.021)***	0.207 (0.019)***
250 to 499	0.312 (0.017)***	0.314 (0.017)***	0.267 (0.016)***
500 and more	0.390 (0.020)***	0.392 (0.020)***	0.382 (0.028)***

Table 3: (Continued)

	All workers	All but agricultural workers	All but community and agricultural workers
Education			
No education	0.125 (0.004)***	0.127 (0.007)***	0.129 (0.007)***
Primary school	0.118 (0.014)***	0.125 (0.012)***	0.130 (0.014)***
Secondary school	0.166 (0.012)***	0.169 (0.008)***	0.169 (0.010)***
High school	0.305 (0.014)***	0.312 (0.013)***	0.285 (0.013)***
University	0.620 (0.019)***	0.627 (0.017)***	0.605 (0.029)***
Industry			
Agriculture	-0.255 (0.026)***		
Mining	0.079 (0.064)	0.080 (0.064)	0.191 (0.055)***
Manufacturing	-0.179 (0.030)***	-0.176 (0.029)***	-0.092 (0.005)***
Energy	0.138 (0.025)***	0.138 (0.024)***	0.260 (0.040)***
Construction	-0.013 (0.020)	-0.016 (0.021)	0.046 (0.021)*
Trade	-0.195 (0.023)***	-0.190 (0.023)***	-0.141 (0.009)***
Transportation	-0.059 (0.022)**	-0.057 (0.022)**	0.025 (0.006)***
Finance	-0.078 (0.029)**	-0.074 (0.029)*	
Occupation			
Professionals	0.017 (0.061)	0.017 (0.061)	-0.097 (0.037)**
Technicians	-0.169 (0.062)**	-0.168 (0.062)**	-0.330 (0.041)***
Clerks	-0.303 (0.063)***	-0.304 (0.062)***	-0.434 (0.050)***
Service workers	-0.400 (0.056)***	-0.400 (0.056)***	-0.524 (0.045)***
Skilled agricultural workers	-0.361 (0.063)***	-0.399 (0.063)***	-0.579 (0.047)***
Craftsmen	-0.328 (0.063)***	-0.327 (0.064)***	-0.497 (0.041)***
Plant operators	-0.343 (0.060)***	-0.341 (0.060)***	-0.513 (0.039)***
Elementary occupations	-0.441 (0.069)***	-0.435 (0.070)***	-0.596 (0.049)***
<i>N</i>	379,512	367,095	256,523
<i>R</i> <sup>2</sup>	0.602	0.599	0.497

Notes: (1) Dependent variable is log real hourly wages. (2) Results are presented for different sample specifications. (3) The numbers in parentheses are robust standard errors clustered for within region and within year correlations and (\*),(\*\*) and (\*\*\*) denote significance at 5%, 1% and 0.1%, respectively. (4) The variable Formality is equal to 1 if the worker is registered to the social security system and 0 otherwise. A positive and significant coefficient estimate for the variable Formality suggests higher wages for formal workers.

Table 4: Determinants of Formality Status

	2005	2006	2007	2008	2009
Age	0.149 (0.00518)***	0.161 (0.00517)***	0.170 (0.00533)***	0.166 (0.00529)***	0.152 (0.00514)***
Age <sup>2</sup>	-0.00214 (0.0000696)***	-0.00232 (0.0000701)***	-0.00251 (0.0000721)***	-0.00247 (0.0000710)***	-0.00220 (0.0000682)***
Gender	-0.279 (0.0201)***	-0.286 (0.0194)***	-0.218 (0.0196)***	-0.210 (0.0200)***	-0.198 (0.0198)***
Experience	0.0810 (0.00368)***	0.0751 (0.00334)***	0.0695 (0.00351)***	0.0752 (0.00376)***	0.0783 (0.00387)***
Experience <sup>2</sup>	-0.00191 (0.000148)***	-0.00164 (0.000130)***	-0.00131 (0.000139)***	-0.00151 (0.000150)***	-0.00162 (0.000153)***
Marital Status					
Single	0.145 (0.0527)**	0.219 (0.0502)***	0.0877 (0.0519)	0.101 (0.0525)	0.0777 (0.0494)
Married	0.355 (0.0495)***	0.416 (0.0475)***	0.270 (0.0490)***	0.258 (0.0494)***	0.200 (0.0458)***
Firm Size					
10 to 24	0.760 (0.0219)***	0.694 (0.0212)***	0.687 (0.0217)***	0.689 (0.0225)***	0.683 (0.0218)***
25 to 49	1.056 (0.0220)***	1.050 (0.0216)***	1.011 (0.0216)***	0.950 (0.0216)***	0.935 (0.0218)***
50 to 249	1.539 (0.0228)***	1.417 (0.0231)***	1.352 (0.0232)***	1.361 (0.0248)***	1.313 (0.0243)***
250 to 499	1.726 (0.0610)***	1.724 (0.0486)***	1.662 (0.0520)***	1.559 (0.0493)***	1.447 (0.0492)***
500 and more	2.008 (0.0723)***	1.882 (0.0525)***	1.873 (0.0606)***	1.878 (0.0594)***	1.639 (0.0559)***
Education					
No education	0.00747 (0.0781)	0.0430 (0.0784)	0.0537 (0.0838)	0.0379 (0.0799)	-1.206 (0.0490)***
Primary school	0.484 (0.0645)***	0.444 (0.0647)***	0.520 (0.0719)***	0.479 (0.0695)***	-0.705 (0.0362)***
Secondary school	0.655 (0.0663)***	0.622 (0.0663)***	0.636 (0.0731)***	0.555 (0.0708)***	-0.595 (0.0365)***
High school	1.002 (0.0664)***	0.964 (0.0664)***	0.992 (0.0732)***	0.948 (0.0708)***	-0.270 (0.0345)***
University	1.295 (0.0728)***	1.323 (0.0725)***	1.302 (0.0791)***	1.232 (0.0771)***	

Table 4: (Continued)

	2005	2006	2007	2008	2009
Occupation					
Professionals	0.357 (0.0610)***	0.348 (0.0578)***	0.270 (0.0629)***	0.401 (0.0658)***	0.410 (0.0682)***
Technicians	0.0549 (0.0521)	0.124 (0.0501)*	0.116 (0.0520)*	0.154 (0.0520)**	0.0236 (0.0517)
Clerks	0.191 (0.0514)***	0.228 (0.0497)***	0.188 (0.0514)***	0.185 (0.0509)***	0.146 (0.0508)**
Service workers	-0.325 (0.0465)***	-0.269 (0.0450)***	-0.263 (0.0468)***	-0.200 (0.0465)***	-0.340 (0.0458)***
Skilled agricultural workers	0.250 (0.162)	0.396 (0.165)*	0.440 (0.149)**	0.512 (0.130)***	0.103 (0.142)
Craftsmen	-0.268 (0.0477)***	-0.129 (0.0465)**	-0.165 (0.0487)***	-0.144 (0.0479)**	-0.190 (0.0477)***
Plant operators	-0.218 (0.0487)***	-0.0935 (0.0473)*	-0.193 (0.0495)***	-0.0925 (0.0493)	-0.149 (0.0483)**
Elementary occupations	-0.207 (0.0478)***	-0.110 (0.0462)*	-0.106 (0.0484)*	-0.0607 (0.0480)	-0.157 (0.0473)***
Number of Observations	70,487	72,964	73,629	75,078	74,937

Notes: (1) Estimated with probit to predict the propensity scores used in semi-parametric estimation for wage gaps. (2) Results are presented for the main sample where workers in the agricultural sector are excluded. (3) The numbers in parentheses are robust standard errors clustered for within region correlations and (\*),(\*\*) and (\*\*\*) denote significance at 5%, 1% and 0.1%, respectively.

Table 5: Formal/Informal Wage Gap by Nuts Regions and Years - Caliper Matching Estimators

	2005	2006	2007	2008	2009
Nuts Region					
1	0.099 (0.020)***	0.043 (0.027)	0.011 (0.032)	0.002 (0.042)	0.135 (0.026)***
2	0.179 (0.113)	0.281 (0.132)*	0.006 (0.201)	0.048 (0.269)	0.368 (0.092)***
3	0.339 (0.160)*	0.109 (0.126)	0.127 (0.148)	0.308 (0.098)**	0.094 (0.091)
4	0.063 (0.065)	-0.004 (0.056)	0.112 (0.058)	0.146 (0.082)	0.201 (0.064)**
5	0.087 (0.087)	-0.084 (0.139)	0.006 (0.125)	0.252 (0.129)	0.204 (0.126)
6	0.239 (0.079)**	0.252 (0.075)***	0.233 (0.055)***	0.159 (0.071)*	0.089 (0.055)
7	0.129 (0.041)**	0.284 (0.099)**	0.053 (0.061)	-0.269 (0.200)	0.000 (0.099)
8	0.274 (0.049)***	0.317 (0.053)***	0.212 (0.061)***	0.278 (0.065)***	0.080 (0.067)
9	0.139 (0.096)	-0.108 (0.129)	0.234 (0.075)**	0.203 (0.051)***	0.198 (0.081)*
10	0.424 (0.136)**	0.496 (0.215)*	0.195 (0.111)	0.358 (0.124)**	0.259 (0.082)**
11	0.247 (0.113)*	0.316 (0.120)**	0.162 (0.120)	0.184 (0.108)	0.214 (0.112)
12	0.042 (0.127)	0.300 (0.071)***	0.321 (0.078)***	0.285 (0.079)***	0.290 (0.078)***
13	0.558 (0.204)**	0.432 (0.130)***	0.394 (0.153)**	0.383 (0.086)***	0.247 (0.119)*
14	0.185 (0.150)	0.435 (0.168)**	0.029 (0.162)	0.149 (0.227)	0.325 (0.145)*
15	0.392 (0.153)*	-0.517 (0.554)	-0.020 (0.190)	0.229 (0.097)*	0.017 (0.110)
16	0.197 (0.097)*	0.015 (0.235)	0.185 (0.161)	0.152 (0.197)	-0.190 (0.205)
17	0.489 (0.365)	0.404 (0.457)	0.250 (0.276)	0.235 (0.158)	0.372 (0.124)**
18	0.357 (0.123)**	0.180 (0.098)	0.284 (0.082)***	0.288 (0.102)**	0.179 (0.092)
19	0.318 (0.177)	0.344 (0.110)**	0.120 (0.130)	0.279 (0.139)*	0.413 (0.098)***
20	0.429 (0.245)	0.221 (0.183)	0.088 (0.287)	0.500 (0.223)*	0.187 (0.286)
21	0.523 (0.270)	0.254 (0.355)	0.298 (0.198)	0.052 (0.207)	0.747 (0.222)***
22	0.242 (0.340)	-0.189 (0.094)*	0.500 (0.243)*	0.186 (0.237)	0.155 (0.122)
23	0.541 (0.273)*	-0.099 (0.181)	0.473 (0.241)*	0.743 (0.302)*	- -
24	0.183 (0.078)*	0.197 (0.138)	0.238 (0.097)*	0.355 (0.131)**	0.391 (0.130)**
25	0.213 (0.161)	0.266 (0.144)	0.238 (0.154)	0.009 (0.210)	0.267 (0.143)
26	0.059 (0.173)	0.290 (0.121)*	0.377 (0.181)*	0.208 (0.153)	0.167 (0.150)
All regions	0.217 (0.021)***	0.155 (0.030)***	0.103 (0.028)***	0.113 (0.037)**	0.214 (0.015)***

Notes: (1) Caliper matching estimators of formal wage gap for 26 regions and 5 years are presented in the table. Positive numbers indicate higher wages for formal workers.  $\delta$  is chosen to be  $10^{-4}$ . (2) Results are presented for the main sample where workers in the agricultural sector are excluded. (3) The numbers in parentheses are robust standard errors and (\*),(\*\*) and (\*\*\*) denote significance at 5%, 1% and 0.1%, respectively. Standard errors for "All regions" are clustered for within region correlations.

Table 6: Formal/Informal Wage Gap by Nuts Regions and Years - Nearest Neighbor Matching Estimators

	2005	2006	2007	2008	2009
Nuts Region					
1	0.095 (0.040)*	-0.014 (0.063)	-0.010 (0.034)	0.019 (0.063)	0.157 (0.034)***
2	0.075 (0.083)	0.162 (0.061)**	0.103 (0.133)	-0.007 (0.212)	0.332 (0.048)***
3	0.274 (0.095)**	0.144 (0.071)*	-0.034 (0.133)	0.422 (0.071)***	0.241 (0.097)*
4	0.011 (0.063)	-0.156 (0.076)*	0.144 (0.059)*	0.177 (0.094)	0.175 (0.057)**
5	0.246 (0.045)***	-0.014 (0.073)	-0.138 (0.086)	0.229 (0.050)***	0.189 (0.088)*
6	0.271 (0.053)***	0.259 (0.077)***	0.244 (0.044)***	0.134 (0.077)	0.197 (0.046)***
7	0.106 (0.051)*	0.369 (0.078)***	0.056 (0.068)	-0.148 (0.120)	0.072 (0.072)
8	0.273 (0.054)***	0.350 (0.047)***	0.312 (0.106)**	0.314 (0.070)***	0.099 (0.050)*
9	0.047 (0.129)	-0.013 (0.100)	0.197 (0.073)**	0.077 (0.073)	0.137 (0.077)
10	0.386 (0.048)***	0.428 (0.197)*	0.128 (0.061)*	0.221 (0.130)	0.275 (0.070)***
11	0.243 (0.065)***	0.336 (0.079)***	0.095 (0.068)	0.223 (0.046)***	0.250 (0.057)***
12	0.159 (0.177)	0.435 (0.073)***	0.318 (0.060)***	0.218 (0.091)*	0.286 (0.039)***
13	0.489 (0.107)***	0.543 (0.100)***	0.366 (0.090)***	0.478 (0.047)***	0.278 (0.076)***
14	0.358 (0.135)**	0.349 (0.103)***	0.182 (0.087)*	0.144 (0.072)*	0.222 (0.114)
15	0.154 (0.098)	-1.043 (0.622)	-0.029 (0.179)	0.283 (0.042)***	0.121 (0.103)
16	0.186 (0.074)*	0.503 (0.186)**	0.076 (0.075)	0.189 (0.056)***	0.191 (0.118)
17	0.394 (0.078)***	0.390 (0.162)*	0.453 (0.132)***	0.215 (0.082)**	0.493 (0.084)***
18	0.171 (0.104)	0.317 (0.087)***	0.258 (0.058)***	0.261 (0.095)**	0.317 (0.100)**
19	0.292 (0.123)*	0.249 (0.074)***	0.226 (0.108)*	0.415 (0.089)***	0.369 (0.051)***
20	0.596 (0.153)***	0.635 (0.128)***	0.276 (0.125)*	0.443 (0.213)*	0.360 (0.066)***
21	0.896 (0.107)***	0.405 (0.247)	0.568 (0.108)***	0.578 (0.083)***	0.385 (0.260)
22	0.344 (0.131)**	0.191 (0.049)***	0.380 (0.056)***	0.403 (0.145)**	0.351 (0.054)***
23	1.111 (0.137)***	0.488 (0.081)***	0.279 (0.165)	0.691 (0.214)**	-
24	0.159 (0.044)***	0.270 (0.053)***	0.266 (0.064)***	0.318 (0.056)***	0.230 (0.082)**
25	0.584 (0.094)***	0.515 (0.083)***	0.421 (0.074)***	0.452 (0.255)	0.663 (0.121)***
26	0.353 (0.103)***	0.669 (0.057)***	0.556 (0.094)***	0.462 (0.071)***	0.569 (0.071)***
All regions	0.225 (0.026)***	0.159 (0.033)***	0.092 (0.033)**	0.110 (0.047)*	0.212 (0.017)***

Notes: (1) Nearest neighbor matching estimators of formal wage gap for 26 regions and 5 years are presented in the table. Positive numbers indicate higher wages for formal workers.  $n$  is chosen to be 1. (2) Results are presented for the main sample where workers in the agricultural sector are excluded. (3) The numbers in parentheses are robust standard errors and (\*),(\*\*) and (\*\*\*) denote significance at 5%, 1% and 0.1%, respectively. Standard errors for “All regions” are clustered for within region correlations.

Table 7: Comparison of Parametric and Semiparametric Estimators of Informal Wage Gap by Gender

Parametric Estimation		
Male	0.188 (0.032)***	
Female	0.199 (0.048)***	
Semiparametric Estimation		
	Caliper	Nearest Neighbor
Male	0.155 (0.019)***	0.154 (0.019)***
Female	0.251 (0.035)***	0.255 (0.036)***

Notes:

- (1) Parametric and semiparametric estimators of formal wage gap by gender are presented in the table. Caliper matching and nearest neighbor matching parameters  $\delta$  and  $n$  are chosen to be  $10^{-4}$  and 1, respectively.
- (2) Results are presented for the main sample where workers in the agricultural sector are excluded.
- (3) The numbers in parentheses are robust standard errors clustered for within region and within year correlations and (\*),(\*\*) and (\*\*\*) denote significance at 5%, 1% and 0.1%, respectively.

Table 8: The Unemployment Elasticity of Real Hourly Wages By Formality/Informality Status

	All workers			All but agricultural workers			All but community and agricultural workers		
	All	Formal	Informal	All	Formal	Informal	All	Formal	Informal
Fixed effects	-0.017 (0.038)	-0.029 (0.035)	-0.022 (0.060)	-0.022 (0.048)	-0.029 (0.035)	-0.023 (0.058)	-0.037 (0.052)	-0.060 (0.037)	-0.048 (0.062)
FE-2SLS	-0.092 (0.091)	-0.042 (0.090)	-0.238 (0.104)**	-0.099 (0.090)	-0.039 (0.090)	-0.249 (0.099)**	-0.159 (0.096)*	-0.132 (0.108)	-0.288 (0.113)**
Obs.	302, 335	217, 577	84, 758	292, 168	216, 239	75, 929	203, 963	140, 032	63, 931

Notes:

- a) See Appendix for sample coverage.
- b) Robust standard errors in parentheses. \*, \*\* and \*\*\* represent significance at 10%, 5% and 1%, respectively.
- c) In FE-2SLS specification, the logarithm of non-agricultural unemployment rate by region in the previous year is used as an instrument for the logarithm of non-agricultural unemployment rate by region at time  $t$ .

Table 9: The Sensitivity of Formality/Informality Wage Gap to Variations in Unemployment Rates

	(1)	(2)	(3)	(4)
$\log U_{rt}$	0.127 (0.060)**	0.138 (0.071)*	0.111 (0.092)	0.132 (0.113)
Region fixed effects	No	No	Yes	Yes
Year fixed effects	No	Yes	No	Yes

Notes:

- a) Dependent variable is obtained with caliper matching. The values of the dependent variable has been shown in Table 5. The explanatory variable is logarithm of non-agricultural unemployment rates. See Appendix for sample coverage.
- b) Robust standard errors in parentheses. \*, \*\* and \*\*\* represent significance at 10%, 5% and 1%, respectively.