I Z A Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15060

Spatial Wage Curves for Formal and Informal Workers in Turkey

Badi H. Baltagi
Yusuf Soner Başkaya

FEBRUARY 2022

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15060

Spatial Wage Curves for Formal and Informal Workers in Turkey

Badi H. Baltagi
Syracuse University and IZA

Yusuf Soner Başkaya
University of Glasgow

FEBRUARY 2022

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Spatial Wage Curves for Formal and Informal Workers in Turkey*

This paper estimates spatial wage curves for formal and informal workers in Turkey using individual level data from the Turkish Household Labor Force Survey (THLFS) provided by TURKSTAT for the period 2008-2014. Unlike previous studies on wage curves for formal and informal workers, we extend the analysis to allow for spatial effects. We also consider household characteristics that would affect the selection into formal employment, informal employment, and non-employment. We find that the spatial wage curve relation holds both for formal and informal workers in Turkey for a variety of specifications. In general, the wages of informal workers are more sensitive to the unemployment rates of the same region and other regions than formal workers. We find that accounting for the selection into formal and informal employment affects the magnitudes but not the significance of the spatial wage curves for the formal and informal workers with the latter always being larger in absolute value than that for formal workers.

JEL Classification: C21, J30, J60

Keywords: spatial wage curve, spatial weights, regional labor markets, informal labor markets

Corresponding author:
Badi H. Baltagi
Department of Economics and Center for Policy Research
Syracuse University
426 Eggers Hall
Syracuse
NY 13244-1020
USA

E-mail: bbaltagi@syr.edu

* We thank Olga Demidova, Murat Kirdar, Semih Tumen, Insan Tunali and participants of the XV World Conference of Spatial Econometrics Association (SEA 2021), virtual, May 26-28, 2021, as well as the participants of the 26th International Panel Data Conference, virtual, July 1, 2021, for their comments and suggestions. We also thank two anonymous referees for their useful comments that helped improve the paper.

1. Introduction

The inverse relation between individual wages and regional employment rates, i.e., *the wage curve*, is regarded as an empirical law in labor economics. The findings in the literature, see Blanchflower and Oswald (1990, 1995), show that such a relationship holds for a wide range of countries and at different time periods.[1] While the early literature simply focused on how the wages for a region covaries with the unemployment rates for that region, more recent studies also focused on how the unemployment rates in the neighboring regions affect these wages. On the economics side, such studies provide insights into whether increases in unemployment rates in a region act as suppressing factor for the wages in the *neighboring regions*, where the neighborhood across geographical regions is defined with respect to whether they are linked via geographical or socio-economic characteristics. Incorporating the spatial dimension of the wage curve relationship is important as the absence of the spatial effects can lead to biased and inconsistent estimates for the relationship between the wages and the unemployment rates of the same region. With this motivation, several studies have estimated spatial wages curves for different countries, such as Longhi, Nijkamp and Poot (2006) and Baltagi, Blien and Wolf (2012) for Germany; Elhorst, Blien and Wolf (2007) for East Germany; Baltagi and Rokicki (2014) for Poland; Ramos, Nicodemo and Sanromá (2015) for Spain, and Karatas (2017) for Turkey to mention a few. The spatial wage curve literature mainly focuses on the spatial aspect of the relationship for the entire population as well as specific demographic groups like age, gender, and education.

One of the key characteristics of labor markets, especially for developing countries, is the considerable degree of informal employment. Despite the large size of the informal labor markets, and the significant need to understand the interplay between informality and earnings, a relatively small number of studies have analyzed wage curves along the formal and informal divide. For example, Ramos et al (2010) analyzed whether the elasticities of own wages with respect to unemployment rates differ across formal and informal workers in Colombia. Baltagi et al (2013) and Baltagi et al (2017) conducted a similar analysis for Turkey and Brazil, respectively. All these papers consistently find that the wages of informal workers are significantly more sensitive to variations in the unemployment rates of the region than wages of their formal counterparts. These findings contribute to the understanding of several policy related issues such as whether wage gaps between formal and informal workers change in response to swings in unemployment rates or whether

[1] See also Card (1995) for a comprehensive survey of the earlier wage curve literature.

informal labor markets play a different role than the formal labor markets in the wage adjustment response to labor market slack. This paper, as far as we know, is the first to add the spatial dimension to the wage curve literature along the formality/informality divide.

We ask whether the differences between the wage curves for informal and formal workers are still observed when we incorporate the spatial effects of the unemployment rates. In doing so, we estimate spatial wage curves for Turkey using individual level data from the Turkish Household Labor Force Survey (THLFS) provided by TURKSTAT for the period 2008-2014. We use NUTS2 level variations in the regional unemployment rate.

Our analysis is important to assess whether the earlier findings on the wage curves for the formal and informal workers are robust to the inclusion of unemployment variations in the neighboring regions. In particular, we empirically estimate whether the surges in the unemployment rates in the regions that are in economic and geographic proximity have disproportional effects on the wages of formal and informal workers. As another novel aspect of our analysis considering the earlier studies on the wage curves along the formality/informality divide, we model the selection into formal employment, informal employment, and non-employment explicitly. We follow a two-step procedure with a multinomial logit model projecting the employment/formality status on household characteristics, such as total household size, number of infants, children, elderly, and active members in the household, which are utilized in the literature on earnings of informal and formal workers, as factors that affect the formality/employment choice but do not have a direct effect on wages.

Finally, we also check the sensitivity of our results with respect to three different spatial weight matrices. In addition to the standard spatial weight matrices based on geographical characteristics such as inverse distances and the contiguity of the NUTS2 level regions, we use another spatial weight matrix based on the migration flows between regions. Studies in Turkey on the interplay between the internal migration and the regional labor market dynamics argue that unemployment in the originating region is one of the key factors affecting internal migration flows. Moreover, these studies also suggest that internal migration is one of the factors feeding the informal employment in the destination regions.[2] These observations motivate us to extend our analysis to include spatial

[2] See Bulutay and Tasti (2004) and Bahar and Bingol (2010).

weights based on migration flows and how unemployment rates in other regions affect the wages of the formal and informal workers.

In summary, we show that the hourly wages of the workers respond both to the variations in the region's own unemployment rate and the unemployment rates of the neighboring regions. We further estimate different spatial wage curves with respect to the informality status of the workers and by gender groups. As in the case for standard wage curves, we find that informal workers experience steeper spatial wage curves compared to formal workers. We also find that the differences in the slope of the wage curves are more pronounced between informal females and informal males, as well as between informal females and formal females. These results provide further evidence for the significant differences in the labor market outcomes of formal and informal workers beyond the earlier results of Baltagi et al (2013).

The rest of the paper is organized as follows: Section 2 reviews the previous literature, while Section 3 presents the data, the model and the spatial weight matrices used. Section 4 presents the estimation results, while Section 5 concludes.

2. Previous Literature

A particular focus of our study is the informality in the labor markets, which is a stylized characteristic of the labor market in developing countries. As Freeman (2010) emphasizes, prevalence of sizable and persistent informality in developing country labor markets creates a significant motivation for studies aiming at understanding normative and positive implications of informal labor markets. Turkish Statistical Institute reports that share of informal employment have been 48% and 36.75% in Turkey. When we look at the sectoral break-down, the informality rate in the non-agricultural sector was almost 34% in 2005 and 22% in 2013. In contrast, for the agricultural sector, which is characterized by the widespread existence of either self-employed or unpaid-family workers rather than salaried employment, the informality rate was 88% in 2005 and 83% in 2013.

Within the context of the wage curves, Baltagi, Baskaya and Hulagu (2013) analyze how the Turkish wage curve differs across the formality-informality divide. They show that the informal workers in Turkey face a much higher decline in wages in response to an increase in the unemployment rates than their formal counterparts. However, their study do not account for spatial effects.

Karatas (2017) also estimates a spatial wage curve for Turkey, so it is important to state how our study differs from his. Karatas (2017) focuses on whether there is a significant spatial wage curve relationship for Turkey for the entire sample and by subsamples defined on the individual's observable characteristics such as age, gender, education, and experience. In contrast, our motivation in this study stems from the prevalence of significant informality in a developing country labor market such as Turkey. This, we argue, has important implications for the labor market. We focus on whether the Turkish spatial wage curve differs across formal and informal labor markets. This is important for understanding several policy relevant questions such as whether earning gaps between formal and informal workers widen with higher unemployment rates in the local labor markets and the neighboring regions. Analysis of such a potential difference is also important for understanding the potential heterogeneities across informal and formal labor markets in the adjustments to unemployment shocks. Another important difference that distinguishes our study from Karatas (2017), is that we use non-agricultural unemployment rates in Turkey while his study uses the overall unemployment rates. It is well-documented that using the overall unemployment rates can be misleading (see Bulutay and Tasti (2002) and Gursel and Acar (2012)). In fact, the overall unemployment rates include unpaid family workers in the agricultural sector. These workers, mostly females, have a marginal attachment to the activities in the labor market, as their employment behavior is mainly determined by non-wage characteristics such as social norms. In addition, there is significant heterogeneity across NUTS2 regions in terms of the size of the agriculture sector and the degree of unpaid employment. This is well-documented for labor markets in Turkey and similar developing countries, leading to some reservations in using the overall unemployment rates.[3] We also show that, unlike the results using non-agricultural unemployment rates, some estimation results with the overall unemployment rates may be difficult to reconcile with theory, intuition, or the stylized labor market facts in Turkey. These results are not shown here to save space but are available upon request from the authors. Finally, the time periods differed across Karatas (2017) and our study due to data availability on some variables.

[3] For example, see Bulutay and Tasti (2002) and Gursel and Acar (2012) to mention a few. Also, the earlier analysis by Baltagi et al (2012, 2013), used non-agricultural unemployment rates deliberately due to this issue.

Finally, we consider the factors that may affect selection into non-employment, informal employment, and formal employment in our wage curve analysis. From this perspective, our analysis differs from earlier studies on the standard and spatial wage curves in Turkey such as Baltagi et al (2012, 2013) or Karatas (2017). This point also constitutes a significant difference from the other studies on the wage curves for formal and informal workers such as Ramos et al (2010), Baltagi et al (2013) and Baltagi et al (2017).

3. Data and Model

3.1 Data

We use individual level data from the Turkish Household Labor Force Survey (THLFS) provided by TURKSTAT for the 2005-2014 period. The survey is conducted with households at the province level including both rural and urban settlements, compiled according to 26 NUTS2 level regions nationwide. Economic activities in the survey are classified according to International Classification of Economic Activities in the European Union (NACE) Rev 2. The sample is restricted to non-institutional working age population, which represents individuals above 15 years of age who are not residents of an institution such as dormitories of universities, orphanage, rest homes for the elderly, special hospitals, prisons, and military barracks. The other key data set, interregional migration statistics which exploits the migration dynamics between 26 NUTS2 level regions, are provided by TURKSTAT since 2008. Combining the two data sets, our sample covers the period between 2008 and 2014.

As we are interested in the estimation of wage curves, we focus on the informal salaried workers. Therefore, we exclude the unpaid family workers, for whom there is no wage and salary. We also exclude the self-employed workers from the analysis following a large number of papers in Turkey.[4] This choice is also guided by some measurement issues for earnings of self-employed individuals. As Badaoui et al (2008, 2010) argue, the reported earnings of self-employed workers may have huge measurement error. For example, the earnings of self-employed individuals potentially incorporate returns to taking risk, which should be regarded as independent from wages. Badaoui et al (2008, 2010) also argue that the social security registration status in one's own enterprise is likely to be less constrained. Therefore, informal employment for a self-employed individuals may be affected by

[4] See Baltagi, Baskaya and Hulagu (2013), Balkan and Tumen (2016) and Ceritoglu et al (2017) among others.

different factors, rather than being unable to get a formal sector job. After excluding unpaid family workers, self-employed workers, employers and missing observations, the sample includes 623,519 observations, of whom 492,293 are formal workers and 131,226 are informal workers.

The dependent variable is the log of hourly real wages which is calculated by dividing the sum of monthly after-tax wage income and additional monthly earnings such as premiums and bonuses by the total hours worked in that month. Following Blanchflower and Oswald (1995) who suggest using regional price indexes to measure real wages in different regions, we deflate the wage series by the regional price indexes provided by TURKSTAT at the NUTS2 level. The regional non-agricultural unemployment rates are from Gursel and Acar (2012) whose calculations are based on the data from TURKSTAT.

3.2 The Model

Following Blanchflower and Oswald (1990, 1995), the model for estimating the standard wage curve can be written as:

$$logW_{irt} = \alpha + \beta\ logU_{rt} + X'_{irt}\gamma + \mu_r + \varphi_t + \vartheta_{irt} \qquad (1)$$

where $logW_{irt}$ is the log of real hourly wages of individual *i* in region *r,* at time *t*. $logU_{rt}$ is log of the non-agricultural unemployment rate in region *r,* at time *t*. X'_{irt} is the matrix of control variables representing the individual characteristics such as age, gender, marital status, education level, social security registration, individual's tenure, employment location, industry classification, occupational group, permanency of employment, firm size and employment position. μ_r is the region fixed effect, φ_t is the time fixed effect and ϑ_{irt} is the error term. While the standard wage curve model controls for region-specific effects, it ignores the spatial dimension of the labor market. Considering the extended mobility of labor across regions over time, the real hourly wages depend not only on the local unemployment rates of a particular region but also on the unemployment rates of neighboring regions.

The omission of spillovers from neighboring labor markets can lead to biased and inconsistent estimates for the wage curves. To account for that, the standard wage curve equation is augmented with a spatial term accounting for the weighted unemployment rates of neighboring regions:

$$logW_{irt} = \alpha + \beta\ logU_{rt} + \theta \sum_{j\neq r}^{J} \omega_{rj} logU_{jt} + X'_{irt}\gamma + \mu_r + \varphi_t + \vartheta_{irt} \qquad (2)$$

where $\sum_{j\neq r}^{J}\omega_{rj}logU_{jt}$ is a measure of non-agricultural unemployment rates in other regions, constructed using spatial weighting matrices described in detail in the next section.

In this specification, β correspond to the elasticity of real hourly wages in region *r* with respect to own unemployment rates, while the elasticity of real hourly wages in region *r* with respect to other regions' unemployment rates depend on the magnitude of θ and the weight matrix. Keeping other factors constant, negative values of β suggests that a rise in the unemployment rate of own region r can suppress individual wages in region *r*. Likewise, the negative value of θ suggests that a rise in the weighted unemployment rate of neighboring regions can suppress individual wages in region *r*.

Our particular focus in this paper is how spatial wage curves differ for the formal and informal workers. Relatively large size of informal labor markets is one of the key characteristics of the labor markets in developing countries. While earlier studies on understanding how wages differ across formal and informal labor markets almost entirely focused on the earning gaps, incomparably small number of studies have analyzed whether wages of formal and informal workers have different degree of sensitivity to unemployment variations. In particular, the analyses by Ramos et al (2010) for Colombia, Baltagi et al (2013) for Turkey and Baltagi et al (2017) for Brazil are examples of this relatively recent line of work. On the other hand, one of the novel aspects of our study is that we contribute to this literature by analyzing whether formal and informal workers have different spatial wage curve relationships by estimating Equation (3) separately for formal and informal workers:

$$logW_{irt} = \alpha^{S} + \beta^{S}\, logU_{rt} + \theta^{S}\sum_{j\neq r}^{J}\omega_{rj}logU_{jt} + X_{irt}'\gamma^{S} + \mu_{r}^{S} + \varphi_{t}^{S} + \vartheta^{S}{}_{irt} \qquad (3)$$

where *S=f, inf* denote the employment status of the individual taking "*f*" for formal workers and "*inf*" for informal workers. In this set up, $|\beta^{inf}| > |\beta^{f}|$ implies that informal workers' wages are more sensitive than formal workers' wages to variations in local unemployment rates. Also, for the same weight matrix, $|\theta^{inf}| > |\theta^{f}|$ implies that informal workers have higher wage elasticity with respect to other neighboring regions' unemployment rates than formal workers.

Following a long list of studies in the literature, including those focusing on Turkey, we use the definition based on social security registration status of the worker.[5] In our analysis, the worker is regarded as a formal worker if he/she is registered in the compulsory social security system at his or her current job, and as an informal worker if not covered by social security system.[6]

In our basic analysis, we estimate spatial wage curves for formal and informal workers separately. On the other hand, one can argue that the selection into formal and informal markets is not exogenous as it can depend on observable and unobservable characteristics of the workers. If such characteristics are also relevant for unemployment sensitivity of the wages of formal and informal workers, the estimates disregarding the selection issue can potentially suffer biases due to endogeneity of the formality status. In order to account for this, we estimate the model using a two-step algorithm, where the first step uses a multinomial logit model to estimate how formality and employment status depends on household characteristics which directly affect the employment and formality outcomes but do not have a direct effect on wages, and the second step estimates spatial wage curves after including the bias correction factors obtained in the first step. We provide details and results of this exercise in Section 4.5.

3.3 Construction of Spatial Weight Matrices

In order to capture the spatial characteristics of the Turkish wage curve, we construct three different spatial weight matrices. The first two are based on geographical proximity and contiguity, respectively. The third one is based on interregional migration rates between 26 NUTS2 regions.[7]

3.3.a. Distance Based Spatial Weight Matrix

A common approach followed in the spatial wage curve estimation literature is to use spatial weight matrices reflecting information on geographical proximity and/or contiguity of the regions. To compare our results with this baseline, we also followed this practice and used inverse distance based and contiguity-based matrices. This approach is similar to earlier studies in the spatial wage curve literature by Longhi, Nijkamp and Poot, (2006) and Baltagi, Blien and Wolf (2012) for Western Germany, Baltagi and Rokicki (2014) for Poland, Ramos, Nicodemo and Sanromá (2015) for Spain

[5] For the examples of measurement of the informality using the compulsory social security enrollment status, see Maloney (2004) and Ramos, Duque and Surinach (2010). For Turkey, see the OECD study by Jutting and De Laiglesia (2009), and Baltagi, Baskaya and Hulagu (2013), Balkan and Tumen (2016) and Ceritoglu et al (2017) among others

[6] The compulsory social security registration in Turkey provides workers with protection mainly in the form of health care and pensions. The contributions to social security are paid by both workers and employers.

[7] Starting from 2008, TURKSTAT provides the number of people migrating between regions in Turkey on an annual basis. Interregional migration rates are computed as the ratio of the number of migrants from region j to region i over the total number of migrants to region i in a particular year.

and Baltagi, Rokicki and Souza (2017) for Brazil. This weight matrix, denoted by Ω_1, is a $26X26$ matrix, where the typical element measures the inverse distances between two centroids for each 26 NUTS2 regions. More formally, Ω_1 is given by:

$$\omega_{rs} = \begin{cases} (d_{rs})^{-1} & if\ r \neq s \\ 0 & if\ r = s \end{cases} \tag{4}$$

where d_{rs} is the air distance between regions *r* and *s*. For each 26 NUTS2 regions, we take the major cities within the region as the centroids. Finally, Ω_1 is row normalized.

3.3.b. Contiguity Based Spatial Weight Matrix

The other alternative exogenous spatial weight matrix, a $26X26$ time invariant matrix denoted by Ω_2, is constructed by considering the spillover effects of regions that share a common border:

$$\omega_{rs} = \begin{cases} c_{rs} & if\ r \neq s \\ 0 & if\ r = s \end{cases} \tag{5}$$

where, c_{rs} is equal to 0 for non-contagious regions and 1 for the contagious regions *r* and *s.* Finally, Ω_2, is row normalized. Therefore, this matrix allows for spatial interaction between two regions if and only if they are contiguous to each other. However, one disadvantage of this approach is that two regions that do not share a geographical border may still be in proximity in terms of economic and social characteristics, which is not captured with this matrix.

3.3.c. Migration Based Spatial Weight Matrix

The spatial matrix Ω_3 is constructed considering the migration flows between NUTS2 regions. In particular, the matrix can be defined as follows:

$$\omega_{rs} = \begin{cases} m_{rst} & if\ r \neq s \\ 0 & if\ r = s \end{cases} \tag{6}$$

where m_{rst} is the typical element of the interregional migration matrix representing the ratio of migration flows into the region *r* from a specific region *s* to total migration flows into the region *r,* at time *t*. This matrix accounts for social and economic factors in determining the spatial mobility of workers due to migration flows, which may have a particular relevance for the spatial unemployment effects on the wages of informal workers. In particular, as explained above, studies in the literature on labor markets in Turkey, such as Bulutay and Tasti (2002) and Bahar and Bingol (2010), argue that there is a relationship between internal migration flows and the informal labor markets in the destination regions. In that sense, the specification with migration based spatial weights can represent the wage pressures experienced by informal workers due to increase in unemployment rates in other regions.

Compared to the previous spatial weight matrices which depend on the geographic characteristics of the regions, one may argue that migration based weights might be affected by the wage dynamics across regions. As Rincke (2010) argues, there is always a trade-off between using exogenous but economically less relevant geography based spatial weights, and economically more relevant spatial weights, which may be criticized to be endogenous under some circumstances.[8] While acknowledging this possibility, we believe that our reliance on the individual level data on earnings and NUTS2 level migration flows, along with controlling for the region fixed effects, relieves this concern to a great extent. Moreover, the literature on mass migration across regions within a country, point to the role of push factors such as the surplus labor in the agricultural sector and socioeconomic factors such as the availability of education services, which decreases the sensitivity of migration flows to changes in wages in the short run.

4 Estimation Results

4.1 Standard Wage Curve Revisited

Table 1 presents estimation results for the standard wage curve for Turkey given in Equation (1) which does not allow for spatial spillovers among regional labor markets. The only difference from the earlier estimates is that the time-period here is 2008-2014 which is dictated by the availability of the regional migration flow data. Columns 1 and 2 differ from each other in terms of whether the

[8] Other examples of papers that use spatial weights based on socio-economic characteristics of these regions include Cohen and Morrison Paul (2004) who estimated a spatial cost-function with the elements of the weighting matrix depending on the share of the value of goods shipped from a state, Parent and Lesage (2008) who analyzed knowledge spillovers using a spatial weighting matrix based on a technological proximity index, Conley and Topa (2002) who used a socio-economic distance based on social networks to analyze the spatial patterns of unemployment in Chicago, and Figlio, Kolpin and Reid (1999) who use interstate migration flows to estimate how welfare policies of states depend on neighbor states' welfare policies. In addition, Baltagi, Blien and Wolf (2012) estimate spatial wage curves for Germany using the proportion of commuters from one region to the other, to mention a few.

own unemployment rate is treated as exogenous or not. Table 1, column 2 uses the lagged unemployment rate as an instrument for the contemporaneous unemployment rate.[9] We replicate the earlier results, ignoring spatial effects, finding a significant wage curve relationship for Turkey with unemployment elasticity -0.085 with respect to real hourly wages.

4.2 Spatial Wage Curves

Table 2 presents the estimates for the spatial wage curves for Turkey. Column 1 presents the estimation results based on the contiguity weight matrix. These results suggest that the elasticity of hourly wages with respect to region's own unemployment rates is -0.098. We also find that hourly wages respond significantly to variations in the weighted neighboring regions unemployment rate with an estimate of θ of -0.129.

Column 2 of Table 2 presents the results based on the inverse distance between centroids of the regions. As in the case of migration-based weights, we find a significant elasticity with respect to the region's own unemployment rate of -0.082 as well as a significant estimate of θ with respect to the weighted neighbors' unemployment rates of -0.062.

Finally, we estimate the specification using spatial weights based on the migration flows. For this weight matrix, the estimates for the elasticity of real wages with respect to own region's unemployment rate is -0.089, while the estimate of θ with respect to the weighted neighbors' unemployment rates is -0.31. Both estimates are statistically significant.

4.3 Spatial Wage Curves along the Formality-Informality Divide

Table 3 gives the estimation results for our spatial wage curves based on the formality-informality divide for different weighting matrices. We obtain a number of results. First, in all specifications, we find that the wages of informal workers are more sensitive to the variations in the unemployment rates than the wages of formal workers. This result is consistent with Baltagi et al (2013), who also found steeper wage curves for informal workers in Turkey. Using the migration weight matrix in column 6 of Table 3, the elasticity estimate for wages with respect to own region's unemployment rate is -0.039 for formal workers, and -0.138 in column 5 of Table 3 for informal workers. The magnitudes of these elasticities are (-0.034 and -0.13) when using the inverse distance weight

[9] See Blanchflower and Oswald (1995) for suggesting the choice of lagged unemployment rate as an instrument.

matrix.[10] Regarding the regional spillovers, we find significant effects of other regions' unemployment rates on the wages of formal and informal workers. Using the spatial weights based on migration flows, the magnitude of this spillovers estimate of θ is more than double for informal than formal workers' wages -0.458 compared to -0.172. For the inverse distance matrix, the comparable numbers are -0.105 and -0.041, all significant. However, using the spatial weights based on the contiguity weight matrix, we find that the regional spillovers are insignificant for formal workers.

These results provide novel insights into how labor markets work in Turkey. First of all, these results suggest that the informal employment plays an important role in the labor market adjustment to macroeconomic shocks resulting in a surge in the unemployment rates. If we interpret the estimated wages elasticities as a measure of labor market flexibility, our results indicate that the wages of informal workers respond to the variations in unemployment rates more than the wages of formal workers. The evidence of such a flexibility has previously been documented by Baltagi, Baskaya and Hulagu (2013). However, as a totally new and interesting result, this paper documents that such labor market flexibility associated with informal employment has a spatial element. In particular, the wages of informal workers in a region shows a larger response to the unemployment rates in the neighboring regions than the formal workers.

As a possible mechanism, this suggests that following an increase in unemployment rates in a region, the informal labor markets in the regions with geographic or socioeconomic connection may provide employment opportunities for the workers who lost their jobs following the shock. The reason why this is more visible for the informal workers can be linked to the fact that the skill level required for the informal jobs are significantly lower than the formal jobs. Moreover, prevalent noncompliance with minimum wage laws in informal markets can be another reason why we observe much larger downward wage flexibility for the informal workers than the formal workers following a surge in the unemployment rates. While our analysis does not provide solid proof for the mechanism, we can speculate that the migration flows of informal workers may be a possible mechanism for this effect. This observation is consistent with the findings based on the spatial matrix using migration flows. Furthermore, the results using contagious regions and inverse-distance matrices can still be consistent with interregional migration flows being one of the channels for the results, as geographic distances play an important role in migration-gravity type models

[10] See columns 3 and 4 of Table 3.

Following the interpretation of these results as a measure of labor market flexibility that arises mainly due to informal employment along with a spatial dimension, the current downward trend in informal employment in many developing countries, provides a further motivation for the need for the policy measures to increase labor market flexibility within formal labor markets. From a traditional perspective, the results suggest that measures that would provide social security along with reduced labor market rigidities is desirable. However, our results also highlight the important of policies to increase spatial mobility within a formal labor market framework. In particular, along with the desired path of decreasing the role of informal markets and promoting formal labor markets, the potential role of spatial flexibility provided by the informal labor markets should be substituted with the spatial flexibility provided by the formal labor markets. These can be achieved via various policies targeting lower regional relocation costs. For example, the policies aiming at lower job search costs or lower social and economic adaptation costs can be necessary for a smooth transition from high degree of informality to high degree of formality in labor markets while retaining a similar degree of labor market flexibility.

4.4. Informal and Formal Spatial Wage Curves Along the Gender Divide

Earlier studies estimating standard wage curves for the Turkish economy find that wages of females are more sensitive to variations in the unemployment rates than their male counterparts. We therefore repeat our analysis by gender, see Table 4 for males and Table 5 for females. We also present the spatial wage curve elasticities for formal/informal status, by gender.

For the migration weight matrix, ignoring the formal/informal status one finds the following results for females and males wage curve elasticities with respect to own region's unemployment rate presented in column 7 of Tables 4 and 5. These are -0.091 for males and -0.067 for females, all statistically significant. The corresponding estimates for θ are -0.297 for males and -0.323 for females, all statistically significant. In contrast, when one accounts for the formal/informal status, the differences between these wage curve estimates by gender yield drastically different results. For the migration weight matrix, these estimates for formal males given in column 8 of Table 4 are almost half -0.049 and -0.174, both statistically significant. In contrast, these estimates for informal females given in column 9 of Table 5 are much higher in absolute value -0.267 and -0.696, both statistically significant. These estimates exhibit the same phenomena for the inverse distance weighting matrix given in columns 5 and 6 of Table 5, where ignoring the formal/informal status of females yields -0.057 and -0.123 estimates, while the corresponding informal female estimates are -0.243 and -0.221, all statistically significant. Comparing Tables 4 and 5, our findings indicate that, regardless of the specification of the spatial weight matrix, the wages of informal female workers have much higher sensitivity to unemployment rates in neighboring regions than informal male workers. In particular, the estimates of wages with respect to other region's unemployment rates for informal female workers is 2-3 times higher in absolute value than that for informal male workers. For the contiguity weight matrix, for both formal and informal males, the estimates of wages with respect to neighboring regions' unemployment rates is not statistically different from zero, see columns 2 and 3 in Table 4. On the other hand, for the migration flows and the inverse distance weighting matrices, we find significant estimates for the weighted unemployment rate in neighboring regions for formal male workers. For female formal workers, we find that the elasticity of wage with respect to own regions' unemployment rate is insignificant across all weight matrix specifications. However, the estimates of wage for this formal female worker with respect to other regions' unemployment rate is significant, except for the specification using the contiguity weight matrix.

4.5. Accounting for Selection into Formality, Informality and Non-employment

In this section, we estimate the spatial wage curves using a two-step procedure where we allow for the selection into the employment status consisting of formal employment, informal employment and the non-employment to be determined by variables that matter for the employment status but do not affect the wages directly. Following Gunther and Launov (2012), we use number of infants, number of children, number of elderly, household size and the number of individuals who have active participation in the labor market in the selection equation.[11] Since the employment status is made up of three distinct states, the selection equation is modeled as a multinomial logit, as opposed to univariate probit as in the standard selection model due to Heckman (1979). The second step involves estimating the spatial wage curve equation with selectivity correction by including in the

[11] One minor change that we had to make with this approach is that we have to conduct all of these analyses with 2008-2013 sample rather than 2008-2014 sample we used before. This is because Turkstat stopped providing information on households younger than 15 years of age in THLFS in 2014 and therefore we cannot construct the 2014 values of the number of infants and number of children variables for the selection equation. For the sake of saving space, we keep the sample for the analysis presented in Table 1-5 as 2008-2014. However, these results are effectively the same if we use between 2008-2013 sample instead of 2008-2014 sample.

regression the inverse Mills ratios denoted by m0 and m1, which are consistent estimators of the conditional expected values of the residuals derived from the selection equation estimated in the first step.[12]

Table 6 shows the result of the selection into employment status for the entire sample, males and females. In all samples, we find that the selection into employment and formality status depend significantly on the number of infants, children, elderly, household size and the number of individuals who have an active participation in the labor market.

Tables 7-9 present the spatial wage curve relationships for informal and formal workers after accounting for selection into employment status. Our results are provided for all three spatial weight matrices and for all individuals, males and females. Before going into the specific findings with respect to the wage curves, it is important to emphasize that the inverse Mills Ratios, denoted by m0 and m1, are significant in the second stage regressions for formal workers and insignificant for informal workers in Table 7. The same results hold for males in Table 8 but things are different for females in Table 9. The significance of these terms suggests that the selection into employment and informality status is not random.

The results in Tables 7 present the spatial wage curve estimates for formal and informal workers for entire sample for different specifications concerning the weight matrices. In short, we find that the elasticity of wages with respect of the unemployment rate in individuals' own region's is around -0.07 to -0.08 for the formal workers with standard errors around 0.006, indicating a significance at the 1 percent significance level. For the informal workers, these figures are around -0.11 to -0.12 depending on the weight matrix with standard errors around 0.011. These figures also suggest that the elasticities faced by the informal workers are significantly higher (in absolute values) than that for the formal workers. When it comes to the estimates of wages with respect to the unemployment rates in the neighboring regions, we find that they are all significantly different from zero at the 1 percent significance level. However, the magnitudes of these estimates differ with respect to different weight matrices. For example, the estimates for informal workers with respect to

[12] For the estimation, we use selmlog command in Stata, based on the Bourguignon, Fournier and Gurgand (2007). The results that we report are based on the bias correction method due to Dubin and McFadden (1984). However, our results are very similar when we alternatively use the bias correction methods due to Lee (1983) or Dahl (2002). Finally, one of the underlying assumptions of the multinomial logit model is the independence of irrelevant alternatives (IIA). The Monte Carlo experiments conducted by Bourguignon et al (2007) shows that selection bias correction based on the multinomial logit model can provide a fairly good correction for the outcome equation, even when the IIA assumption does not hold.

neighboring regions' unemployment rates vary between -0.10 and -0.43 depending on the weight matrix. In contrast, the estimates for the formal workers with respect to neighboring regions' unemployment rates vary between -0.06 and -0.24. For all weight matrices, the estimates for the informal workers indicate a higher sensitivity of wages to unemployment rate variations in other regions compared to the wages of formal workers. The magnitudes of these estimates are different from the results in Table 3, where we did not correct for selection into employment status. The results in Table 7, are all statistically significant and all indicate that informal workers have higher estimates in absolute value than their formal counterparts.

Table 8 gives the results for the spatial wage curves for males after accounting for selection issues. Compared to Table 4 where we did not correct for selection into employment status, the magnitudes are different, some smaller in absolute value and some larger, but they are all statistically significant.

Table 9 gives the results for the spatial wage curves for females after accounting for selection issues. Compared to Table 5 where we did not correct for selection into employment status, the magnitudes are again different, and a few switch from being significant to insignificance and vice versa.

In short, controlling for selection affects the magnitudes but not the significance of the spatial wage curves for formal and informal workers. We consistently find that the informal workers always have larger estimates in absolute value than that for formal workers.

5 Conclusion

This paper extends the literature on spatial wage curves and the informal labor markets by asking whether the spatial wage curves differ for the formal and informal workers for Turkey. Our findings contribute to the wage curve literature in general and specifically for the informal and formal workers in a number of dimensions. We first show that spatial wage curve relationship holds both for formal and informal workers. This is a consistent finding across all specifications especially after we control for the selection into formal employment, informal employment, and non-employment, which is a dimension not incorporated in earlier papers on wage curves for formal and informal workers. Second, for all specifications, we find that sensitivity of wages with respect to the unemployment rates in the individual's own regions is significantly higher for informal workers than formal workers, which is in line with earlier studies that did not have the spatial dimension. Third, we argue that accounting for selection into formal employment, informal employment and the non-employment can matter for the estimated wage elasticities and the comparison between the spatial wage curves of the formal and informal workers.

Finally, our results highlight the role of informal labor markets in absorbing not only the local labor market shocks but also the shocks in the neighboring regions. These also suggest that long-term objectives in developing countries, such as Turkey, in favor of replacing informal labor markets with formal structures should be supported with policies that aim at strengthening both local labor market flexibility and spatial mobility of formal workers.

References

Bahar, D and F.K.Bingol (2010), “Turkiye’de Ic Gc Hareketlerinin Istihdam ve Isgucu Piyasalarina Etkileri,” Suleyman Demirel Universitesi IIBF Dergisi, 15 (2), p. 43-61.

Balkan B. and Tumen S. (2016) "Firm-size wage gaps along the formal-informal divide: Theory and evidence." Industrial Relations, 55(2), 235-266.

Baltagi, B.H., Baskaya, Y.S., Hulagu, T., (2012), “The Turkish wage curve: evidence from household labor force survey,” Economic Letters. 114, 128-131.

Baltagi, B.H., Baskaya, Y.S., Hulagu, T., (2013), “How different are the wage curves for formal and informal workers? evidence from Turkey,” Papers in Regional Science, 92, 271-283.

Baltagi, B.H., Blien, U., Wolf, K., (2012), “A dynamic spatial panel data approach to the German wage curve,” Economic Modelling, 29, 12-21.

Baltagi, B.H., Rokicki, B., (2014), “The spatial Polish wage curve with gender effects: evidence from the Polish labor survey,” Regional Science and Urban Economics, 49, 36-47.

Baltagi B.H., B. Rokicki and K.B Souza, (2017), "The Brazilian wage curve: new evidence from the National Household Survey," Empirical Economics, Springer, 53(1), 267-286.

Blanchflower, D.G., Oswald, A.J., (1990), “The wage curve,” Scandinavian Journal Economics, 92, 215-235.

Blanchflower, D.G., Oswald, A.J, (1995), “An introduction to the wage curve,” The Journal of Economic Perspectives. 9, 153–167.

Bourguignon, F., Fournier, M. and Gurgand, M., (2007), “Selection Bias Corrections Based on the Multinomial Logit Model: Monte Carlo Comparisons,” Journal of Economic Surveys, 21, 174-205

Bulutay, T. and Tasti, E. (2004). “Informal Sector in the Turkish Labor Market,” Turkish Economic Association, Working Papers.

Card, D., (1995), “The wage curve: a review,” Journal of Economic Literature, 33, 285–299.

Ceritoglu, E., Gurcihan-Yunculer, H., Torun, H. and Tumen S., (2017), “The impact of Syrian refugees on natives’ labor market outcomes in Turkey: evidence from a quasi-experimental design.” IZA Journal of Labor Policy, 6, 5.

Cohen, J.P., Morrison Paul, C.J., (2004), “Public infrastructure investment, interstate spatial spillovers, and manufacturing costs,” The Review of Economics and Statistics, 86, 551-560.

Conley, T., Topa, G., (2002), “Socio-economic distance and spatial patterns in unemployment,” Journal of Applied Econometrics, 17, 303-327.

Dahl G. B., (2003), "Mobility and the Returns to Education: Testing a Roy Model with Multiple Markets", Econometrica, 70, 2367-2420.

Dubin J.A. & McFadden D.L., (1984), "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption", Econometrica, 52, 345-362.

El Badaoui E., Strobl E. and Walsh F., (2008) "Is There an Informal Employment Wage Penalty? Evidence from South Africa," Economic Development and Cultural Change, 56(3), 683-710.

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.

Elhorst, J.P., Blien, U., Wolf, K., (2007), “New evidence on the wage curve: A spatial panel approach,” International Regional Science Review, 30, 173-191.

Figlio, D.N., Kolpin, V.W., Reid, W.E., (1999), “Do states play welfare games?” Journal of Urban Economics, 46, 437-454.

Freeman, R.B., (2010) Labor regulations, unions, and social protection in developing countries: market distortions or efficient institutions? In: Rodrik, D., Rosenzweig, M. (Eds.), Handbook of Development Economics 5, Elsevier B.V. Amsterdam: North Holland, pp. 4657-4702.

Günther, I., and A. Launov, (2012), “Informal employment in developing countries,” Journal of Development Economics, 97(1), 88-98.

Gursel, S. and A. Acar, Aysenur, (2012), “Regional Labor Markets in Turkey”, No 130, Research Briefs, Bahcesehir University, Betam.

Heckman, J., (1979), “Sample Selection Bias as a Specification Error”, Econometrica, 47(1), 153-61.

Ilkkaracan, I., Selim, R., (2003), “The role of unemployment in wage determination: further evidence on the wage curve from Turkey,” Applied Economics, 35, 1589-1598.

Jutting, J., De Laiglesia, J.R., (2009), “Is informal normal? Towards more and better jobs in developing countries,” OECD Development Centre, Paris.

Karatas, M. H., (2017), “The Turkish Spatial Wage Curve,” Econometrics, 5, 37, 1-21

Kleibergen, F. and Paap, R., (2006), “Generalized reduced rank tests using the singular value decomposition” Journal of Econometrics, 133, 97-126.

Lee L.F. (1983), "Generalized Econometric Models with Selectivity", Econometrica, 51, 507-512.

Longhi, S., Nijkamp, P., Poot, J., (2006), “Spatial heterogeneity and the wage curve revisited,” Journal of Regional Science, 46, 707–731.

Maloney, W.F., 2004, “Informality revisited,” World Development, 32, 1159-1178.

Parent, O., LeSage, J., (2008), “Using the variance structure of the conditional autoregressive spatial specification to model knowledge spillovers,” Journal of Applied Econometrics, 23, 235-256.

Ramos, R., Duque, J.C., Surinach, J., (2010), “Is the wage curve formal or informal? Evidence for Colombia,” Economics Letters, 109, 63-65.

Ramos, R., Nicodemo, C., Sanromá, E., (2015), “A spatial panel wage curve for Spain,” Letters in Spatial and Resource Sciences. 8, 125-139

Rincke, J., (2010), “A commuting-based refinement of the contiguity matrix for spatial models, and an application to local police expenditures,” Regional Science and Urban Economics, 40, 324-330.

Turkish Statistical Institute, TURKSTAT, 2015. Household labor force survey data for 2005-2014.

Appendix A: Variable Definitions

- Earnings: Monthly nominal earnings variable is the sum of after tax wage income and additional monthly earnings such as premiums and bonuses. This amount is converted to hourly earnings by dividing it by total hours worked in survey mount. We use regional consumer prices in order to deflate the hourly earnings.
- Age: A set of eleven binary variables, each representing age categories in 5-year intervals.
- Gender: Male=0, female=1.
- Marital status: Married=1, otherwise=0.
- Education: There are there variables associated with individuals' educational background.
  - The variable *educ* stands for the years of completed education.
  - The variable *enrolled* takes the value 1 if the individual is enrolled in a school, and zero otherwise.
  - The variable *attend* takes the value 1 if the enrolled school requires attendance, and zero otherwise.
- Social security registration: Binary variable taking the value 1 if the individual is registered in the social security administration, and zero otherwise.
- The individual's years of tenure: Years of experience in the current job calculated by subtracting the starting year of current job from the survey year.
- Employment location: Urban=1, rural=0.
- Industry classification: A set of eight binary variables categorized according to NACE Rev.2 classification for economic activities. These categories are agriculture, mining, manufacturing, electricity, construction, trade, transportation and finance.
- Occupational group: A set of nine binary variables categorized according to ISCO-88 classification. These categories are legislators, senior officials and managers; professionals; technicians and associate professionals; clerks; service workers and shop and market sales workers; skilled agricultural and fishery workers; craft and related trades workers; plant and machine operators and assemblers; and elementary occupations.
- Permanency of employment: A set of three binary variables representing whether the job is *permanent*, *temporary* or *seasonal*.
- Employment type: Part time=1, full time=0.

- Firm size: A set of six binary variables categorized based on the number of employees in the firm. These categories are less than 10 employees; 10–24 employees; 25–49 employees; 50–249 employees; 250–499 employees; and 500 and more employees.
- Other activity to earn income: Yes=1, no=1.
- Employment status in the same month of last year: This is a binary variable which takes the value 1 if the individual was working the previous year, and zero otherwise.
- Employment position: Binary variable which takes the value 1 if the individual is regular or casual worker, and 0 if the individual is unpaid family worker, self-employed or employer.
- Workplace: Binary variable which takes the value one if the working place is a regular establishment and zero if the working place is an agricultural field, garden, market place, house or some kind of pedlary.

Table A.1. Data

	Process	# of observations		
		Formal	Informal	Total
Step 1	Restricted to non-institutional working age population (individuals above 15 years of age)	554,317	291,807	846,124
Step 2	Excluding unpaid family workers, self-employed workers and employers.	504,341	159,612	663,953
Step 3	Excluding missing variables	492,293	131,226	623,519

Table 1. Standard Wage Curve Estimates 2008-2014

	Fixed Effects	FE-2SLS
$logU_{rt}$	-0.086***	-0.0853***
	(0.0054)	(0.0125)
R^2	0.61	0.61
Kleibergen-Paap F stat.		45317.77
Time Fixed Effects	Yes	Yes
Region Fixed Effects	Yes	Yes
Obs.	623,519	623,519

Notes: * p<0.10; ** p<0.05; *** p<0.01. Robust standard errors in parentheses. In FE-2SLS specification, the logarithm of non-agricultural unemployment rate by region in the previous year is used as instruments. Robust Kleibergen-Paap (Kleibergen and Paap, 2006) Wald rk F statistics suggest that the weak instruments null hypothesis is rejected. Control variables representing the individual characteristics are age, gender, marital status, education level, social security registration, individual's tenure, employment location, industry classification, occupational group, permanency of employment, firm size, and employment position.

Table 2. Spatial Wage Curve Estimates: 2008-2014.

	(1)	(2)	(3)
	FE-2SLS- Weights Based on Contiguity	FE-2SLS- Weights Based on Inverse Distance	FE-2SLS-Weights Based on Migration Flows
$logU_{rt}$	-0.0979***	-0.0817***	-0.0890***
	(0.0158)	(0.0123)	(0.0127)
$\sum_{j\neq r}^{J}(w_j logU_{jt})$	-0.129***	-0.0615***	-0.310***
	(0.0453)	(0.0132)	(0.0341)
R^2	0.62	0.62	0.62
Kleibergen-Paap F stat.	40043.94	26022.86	56004.63
Time FE	Yes	Yes	Yes
Region FE	Yes	Yes	Yes
Obs.	623,519	623,519	623,519

Notes * p<0.10; ** p<0.05; *** p<0.01. Robust standard errors in parentheses.
Control variables representing the individual characteristics are age, gender, marital status, education level, social security registration, individual's tenure, employment location, industry classification, occupational group, permanency of employment, firm size, and employment position.

Table 3. The Spatial Wage Curves Along the Formality-Informality Divide 2008-2014.

	FE-2SLS Weights Based on Contiguity		FE-2SLS Weights Based on Inverse Distance		FE-2SLS Weights Based on Migration	
	(1)	(2)	(3)	(4)	(5)	(6)
	Informal	Formal	Informal	Formal	Informal	Formal
$logU_{rt}$	-0.163***	-0.0376**	-0.130***	-0.0338***	-0.138***	-0.0390***
	(0.0358)	(0.0169)	(0.0297)	(0.0128)	(0.0305)	(0.0134)
$\sum_{j \neq r}^{J}(w_j logU_{jt})$	-0.280***	-0.0254	-0.105***	-0.0405***	-0.458***	-0.172***
	(0.104)	(0.0482)	(0.0282)	(0.0143)	(0.0749)	(0.0371)
R^2	0.321	0.596	0.321	0.596	0.321	0.596
Kleibergen-Paap F stat.	6669.51	36578.12	5401.05	21838.98	10225.76	48765.84
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	131,226	492,293	131,226	492,293	131,226	492,293

Notes: * p<0.10; ** p<0.05; *** p<0.01. Robust standard errors in parentheses.
Control variables representing the individual characteristics are age, gender, marital status, education level, social security registration, individual's tenure, employment location, industry classification, occupational group, permanency of employment, firm size, and employment position.

Table 4. The Spatial Wage Curves Along the Formality-Informality Divide – Males - 2008-2014.

	FE-2SLS Weights Based on Contiguity			FE-2SLS Weights Based on Inverse Distance			FE-2SLS Weights Based on Migration Flows		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All	Formal	Informal	All	Formal	Informal	All	Formal	Informal
$logU_{rt}$	-0.0992***	-0.0480**	-0.103**	-0.0857***	-0.0440***	-0.0869***	-0.0910***	-0.0488***	-0.0882**
	(0.0180)	(0.0194)	(0.0403)	(0.0142)	(0.0150)	(0.0337)	(0.0147)	(0.0156)	(0.0346)
$\sum_{j\neq r}^{J}(w_j logU_{jt})$	-0.107**	-0.0223	-0.0897	-0.0478***	-0.0361**	-0.0611**	-0.297***	-0.174***	-0.341***
	(0.0510)	(0.0549)	(0.117)	(0.0147)	(0.0162)	(0.0307)	(0.0380)	(0.0416)	(0.0835)
R^2	0.595	0.578	0.342	0.595	0.578	0.342	0.596	0.578	0.342
Kleibergen-Paap F stat.	28169.71	2276.52	4744.13	19459.80	16364.44	4101.27	40628.24	35607.13	7447.95
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	468,976	372,027	96,949	468,976	372,027	96,949	468,976	372,027	96,949

Notes:* p<0.10; ** p<0.05; *** p<0.01. Robust standard errors in parentheses.
Control variables representing the individual characteristics are age, marital status, education level, social security registration, individual's tenure, employment location, industry classification, occupational group, permanency of employment, firm size, and employment position.

Table 5. The Spatial Wage Curves Along the Formality-Informality Divide – Females - 2008-2014

	FE-2SLS Weights Based on Contiguity			FE-2SLS Weights Based on Inverse Distance			FE-2SLS Weights Based on Migration Flows		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All	Formal	Informal	All	Formal	Informal	All	Formal	Informal
$logU_{rt}$	-0.0808**	0.0071	-0.309***	-0.0569**	0.0112	-0.243***	-0.0668***	0.00741	-0.267***
	(0.0319)	(0.0329)	(0.0738)	(0.0239)	(0.0241)	(0.0593)	(0.0253)	(0.0257)	(0.0614)
$\sum_{j\neq r}^{J}(w_j logU_{jt})$	-0.213**	-0.0515	-0.579***	-0.123***	-0.0730**	-0.221***	-0.323***	-0.148*	-0.696***
	(0.0962)	(0.0987)	(0.219)	(0.0290)	(0.0295)	(0.0658)	(0.0752)	(0.0795)	(0.162)
R^2	0.669	0.652	0.336	0.669	0.652	0.336	0.669	0.652	0.336
Kleibergen-Paap F stat.	13486.83	11791.86	2276.52	7270.93.	1495.75	5897.93	8471.51	13948.45	3177.04
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	154,543	120,266	34,277	154,543	120,266	34,277	154,543	120,266	34,277

Notes:* p<0.10; ** p<0.05; *** p<0.01. Robust standard errors in parentheses.
Control variables representing the individual characteristics are age, marital status, education level, social security registration, individual's tenure, employment location, industry classification, occupational group, permanency of employment, firm size, and employment position.

Table 6: Selection into Nonemployment, Informal Employment and Formal Employment

Employment Category (Nonemployment (i.e. Employment Status=0) taken as baseline)	*Variables*	All Individuals	Males	Females
***Panel A:** Informal Employment (Employment-Status=1)*	Number of Infants	0.024***	0.254***	-0.259***
		(0.006)	(0.007)	(0.011)
	Number of Children	0.076***	0.128***	0.144***
		(0.004)	(0.005)	(0.007)
	Number of Elderly	0.014**	-0.02**	0.184**
		(0.007)	(0.010)	(0.011)
	Household Size	-0.081***	-0.111***	-0.176***
		(0.002)	(0.003)	(0.005)
	Number of Active Workers	0.282***	0.191***	0.589***
		(0.004)	(0.005)	(0.006)*
	Constant	-0.691***	-0.046***	-1.291***
		(0.006)	(0.008)	(0.010)
***Panel B:** Formal Employment (Employment-Status=1)*	Number of Infants	0.171***	0.435***	-0.306***
		(0.004)	(0.006)	(0.008)
	Number of Children	0.014***	0.172***	-0.289***
		(0.003)	(0.004)	(0.006)
	Number of Elderly	-0.214***	-0.126***	-0.263***
		(0.007)	(0.009)	(0.011)
	Household Size	-0.209***	-0.307***	-0.145***
		(0.002)	(0.003)	(0.003)
	Number of Active Workers	-0.582***	-0.42***	-0.847***
		(0.005)	(0.006)	(0.010)
	Constant	1.049***	1.814***	0.253***
		(0.005)	(0.007)	(0.008)
Number of Observations		931,557	572,675	358,882

Notes: (1) The regressions take non-employment as the baseline category. (2) The coefficients in Panel A report the risk of becoming an informal worker compared to being non-employed when the associated variable increases by 1 unit. The coefficients in Panel B report the risk of the becoming formal worker compared to being non-employed when the associated variable increases by 1 unit. (3) Columns 3, 4 and 5 correspond to all individuals, males and females. (4)The numbers report Robust standard errors, *** and ** indicates significant at 1% and 5% respectively.

Table 7. The Spatial Wage Curves Along the Formality-Informality Divide 2008-2013.

	Weights Based on Inverse Distance Matrix		Weights Based on Contiguity		Weights Based on Inverse Migration Flows	
	(1)	(2)	(3)	(4)	(5)	(6)
	Informal	Formal	Informal	Formal	Informal	Formal
$logU_{rt}$	-0.116***	-0.0761***	-0.109***	-0.0662***	-0.123***	-0.0768***
	(0.0117)	(0.00594)	(0.0110)	(0.00543)	(0.0113)	(0.00562)
$\sum_{j\neq r}^{J}(w_j logU_{jt})$	-0.187***	-0.148***	-0.107***	-0.0567***	-0.434***	-0.238***
	(0.0660)	(0.0301)	(0.0257)	(0.0126)	(0.0610)	(0.0287)
m0	0.005	0.557***	0.002	0.556***	0.002	0.554***
	(0.022)	(0.017)	(0.022)	(0.017)	(0.022)	(0.017)
m1	0.000	-0.484***	0.000	-0.486***	0.000	-0.484***
	(0.013)	(0.016)	(0.013)	(0.016)	(0.013)	(0.016)
R^2	0.299	0.611	0.299	0.611	0.2998	0.6113
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	114,163	413,633	114,163	413,633	114,163	413,633

Notes: * p<0.10; ** p<0.05; *** p<0.01. Robust standard errors in parentheses.
Control variables representing the individual characteristics are age, gender, marital status, education level, social security registration, individual's tenure, employment location, industry classification, occupational group, permanency of employment, firm size and employment position.

Table 8. The Spatial Wage Curves Along the Formality-Informality Divide - Males 2008-2013.

	Weights Based on Inverse Distance		Weights Based on Contiguity		Weights Based on Inverse Migration Flows	
	(1)	(2)	(3)	(4)	(5)	(6)
	Informal	Formal	Informal	Formal	Informal	Formal
$logU_{rt}$	-0.130***	-0.0908***	-0.123***	-0.0806***	-0.137***	-0.0916***
	(0.0125)	(0.00671)	(0.0118)	(0.00616)	(0.0120)	(0.00636)
$\sum_{j\neq r}^{J}(w_j logU_{jt})$	-0.165**	-0.154***	-0.0773***	-0.0542***	-0.411***	-0.253***
	(0.0720)	(0.0342)	(0.0275)	(0.0143)	(0.0660)	(0.0326)
m0	-0.0206	0.262***	-0.0212	0.260***	-0.0222	0.258***
	(0.023)	(0.018)	(0.023)	(0.018)	(0.023)	(0.018)
m1	0.022*	-0.189***	0.0224*	-0.187***	0.022*	-0.185***
	(0.013)	(0.019)	(0.013)	(0.019)	(0.013)	(0.019)
R^2	0.322	0.592	0.322	0.592	0.322	0.592
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	85,170	313,498	85,170	313,498	85,170	313,498

Notes: * p<0.10; ** p<0.05; *** p<0.01. Robust standard errors in parentheses.
Control variables representing the individual characteristics are age, gender, marital status, education level, social security registration, individual's tenure, employment location, industry classification, occupational group, permanency of employment, firm size and employment position.

Table 9. The Spatial Wage Curves Along the Formality-Informality Divide - Females 2008-2013.

	Weights Based on Inverse Distance		Weights Based on Contiguity		Weights Based on Inverse Migration Flows	
	(1)	(2)	(3)	(4)	(5)	(6)
	Informal	Formal	Informal	Formal	Informal	Formal
$logU_{rt}$	-0.0875***	-0.0253**	-0.0919***	-0.0162	-0.105***	-0.0238**
	(0.0280)	(0.0128)	(0.0260)	(0.0114)	(0.0266)	(0.0120)
$\sum_{j\neq r}^{J}(w_j logU_{jt})$	-0.0485	-0.143**	-0.173***	-0.0822***	-0.458***	-0.189***
	(0.146)	(0.0630)	(0.0607)	(0.0266)	(0.136)	(0.0602)
m0	0.086***	0.065***	0.084***	0.063***	0.085***	0.064***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.028)	(0.017)
m1	-0.072***	-0.028	-0.071***	-0.028	-0.072***	-0.028
	(0.018)	(0.017)	(0.018)	(0.017)	(0.018)	(0.017)
R^2	0.321	0.677	0.321	0.677	0.321	0.677
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	28,993	100,135	28,993	100,135	28,993	100,135

Notes: * p<0.10; ** p<0.05; *** p<0.01. Robust standard errors in parentheses.
Control variables representing the individual characteristics are age, gender, marital status, education level, social security registration, individual's tenure, employment location, industry classification, occupational group, permanency of employment, firm size and employment position.