Agglomeration Effects in a Developing Economy: Evidence from Turkey

Cem Özgüzel*

November 26, 2018

Preliminary draft

<Please do not circulate>

Abstract

Spatial inequalities in Turkey are a source of considerable policy concern. In this paper, we estimate agglomeration effects for provinces in Turkey to shed light on the origins of spatial inequality in productivity and provide evidence from a developing country context which literature needs. We use social security data, recently made public at the NUTS-3 level, for 81 provinces of Turkey for the period 2007-2016 and carry out a two-step estimation. We use a variety of panel data techniques and historical instruments to deal with estimation concerns. We estimate an elasticity of labor productivity with respect to the density of 0.038-0.054 which is higher than in developed countries and around the levels observed in developing countries. We also discuss issues that might be limiting the agglomeration effects in Turkey.

Keywords: Local labour markets; Spatial wage disparities; Developing country; Turkey

JEL codes: R12, R23, J31

*I am grateful to Pierre-Philippe Combes for his guidance and support all along this research. I thank Seyhun Orcan Sakalli for sharing his data on historical population statistics, Guney Celbis, Martin Koning, Sandra Poncet and Ivan Savin for useful feedback. All remaining errors are mine

*Paris School of Economics, Université Paris 1 Panthéon-Sorbonne. email: cem.ozguzel@psemail.eu.

1 Introduction

Spatial inequalities in Turkey are a source of considerable policy concern as there are substantial differences between regions in almost every metric (i.e., income, production, life quality, etc.). In this paper, we estimate agglomeration effects for provinces in Turkey to shed light on the sources of spatial inequality in productivity. A better understanding of these factors is crucial for three reasons. First, it would show us which local factors make a given worker more productive. Second, understanding the determinant forces would also make it possible to formulate policies to reduce regional differences. Third, with a population of 78 million, of which 92 percent living in cities makes agglomeration economies a relevant issue for a vast majority of the population. Our findings show that there are indeed substantial local interactions in Turkey that impact workers' productivity.

Agglomeration literature shows that urbanization and development are strongly correlated (Henderson 2010). As countries grow, they undergo structural change, and labor is reallocated from rural agriculture to urban manufacturing and services (Michaels, Rauch and Redding, 2012). Cities enjoy a productivity advantage over rural areas, and this advantage is more significant for larger cities (Krugman, 1991). Although most of the empirical work uses wages as a proxy for productivity, larger cities obtain higher scores on many productivity metrics, such as output per worker or the total factor productivity of firms. Spatial inequality in wages (thus productivity) can be explained through three broad sets of explanations. First, differences in wages across areas could directly reflect spatial variations in the skill composition of the workforce (Rauch, 1993; Moretti, 2004). Knowledge and innovation diffuse faster in areas where the share of skilled workers is higher. Combes et al. (2008) show that indeed, workers sort across employment areas depending on their skills. Moreover, uneven geographic distribution of industries can also impact the local wages. Areas with skill-intensive sectors will have a higher mean wage. For instance Combes et al. (2008) show that skill composition of the labor force accounts for 40 to 50 percent of aggregate spatial wage disparities in France.

The second family of explanations argues that spatial wage differences are caused by variations in local productive and non-productive (non-human) endowments, which could increase the marginal productivity of workers. For instance, geographical features such as a favorable location (like a port or a bridge on a river), a climate more suited to economic activity, or some natural resources could contribute to the productivity of the workers (Roback, 1982; Albouy et al., 2013).

The third strand of explanations contends that some interactions between workers or firms take place locally and lead to productivity gains which are known as Marshallian externalities (Marshall, 1890). Many factors create externalities due to agglomerations, such as denser input-output linkages between buyers and suppliers, facilitation of a better match of workers' skills with firms' due to thicker labor markets, and technological externalities resulting from more intense direct interactions (see Duranton and Puga, 2004 for a review). A key issue is whether these benefits stem from the size of the overall market (urbanization economies) or geographic concentration at the industry level (localization economies).

In this paper, we examine the importance of these three types of explanations on the productivity disparities across Turkish provinces using a "unified framework" à la Combes et al. (2008). Such a unified framework should provide us with a sense of magnitudes on the significance of these explanations which is crucial policy formulation.

Such a broad framework imposes formidable data requirement. To deal appropriately with skills-based explanations, we must control for unobserved worker heterogeneity, which requires a panel of workers. For instance, Combes et al. (2010) use a panel of workers to control for unobserved individual characteristics that may be correlated with location choices. Such individual data are not available in Turkey or most developing countries. To fulfill the objectives of this paper, we use a new data set that has been made public recently which provides aggregated wages at the industry, province and city level.

Our estimation has two steps. In the first step, we assess the importance of industry-specific explanations against those highlighting productivity differences across provinces (i.e., between industry interactions and endowments-based explanations). More specifically, we regress average province-level wages on time-varying variables relating to the local characteristics of the industry, and industry fixed effects and province-year fixed effects to capture local interactions within industries. The province-year fixed effects can be interpreted as local wage indices after controlling for observed and unobserved industry effects.

In the second step, we use the province-year fixed effects estimated in the first step and regress them on several variables capturing local interactions between industries, a set of time dummies, some controls for local endowments and local human capital stocks.

Finally, we use a variety of panel data techniques, and instrumental variables approach to deal with estimation concerns.

The density of local employment (urbanization economies) plays the most important role. The estimate regarding the elasticity of productivity with respect to density is about 0.029-0.038 percent which is lower than one estimated for China (Combes et al., 2015; Chauvin et al. 2016), Colombia (Duranton, 2015) and India (Chauvin et al. 2016) around the ones estimated for Brazil (Chauvin et al. 2016). Domestic market access has elasticity around 0.063-0.090, which is higher than the density in all specifications.

This study is important for several reasons. First, the literature on agglomeration economies mainly comes from developed (Western) country context, and empirical evidence from developing countries is still limited. Moreover, due to data constraints, the limited literature is unable to use the state-of-the-art methods and thus suffers from identification concerns. Combes et al. (2015), for China, Chauvin et al. (2013), for India, and Duranton (2015), for Colombia are the few cases where such concerns have been addressed. This analysis adopts the benchmark approach in the literature in order to provide results that can extend the knowledge base about agglomeration economies and their impact on developing countries.

Second, Turkey suffers from large spatial disparities. These differences have important policy implications as their existence limits the efficient use of national resources, and creates additional problems such as massive internal migration and oversized primate cities. Hence, it is imperative to understand the factors behind these differences so that necessary policies can be formulated to decrease/limit regional inequalities.

Third, as it is commonly used in the literature, we adopt a unified framework to examine the relative importance of the determinants of productivity disparities across Turkey. Hierarchizing these factors is very important to develop necessary policies and to guide future theoretical work on the issue.

The rest of the paper is structured as follows. We first present the context and related work done in Turkey. Then, in Section 3 we propose a general model of spatial wage disparities. In Section 4 we present our data and in Section 5 our results. Section 6 concludes.

2 Context

Turkey has a population of about 78 million over an area of 783 thousand square kilometers(Turkstat, 2016). Large spatial inequalities exist in many dimensions. For instance, Gross Domestic Product (GDP) per capita as of 2014 was $ 12 112, which makes Turkey an upper-income developing country (Turkstat, 2017). However, this wealth is not equally distributed across its regions. While the GDP per capita was $ 19 957 in Istanbul, it was $ 3 880 in Ağrı (Turkstat, 2017). These differences are multiplied even further since population distribution is also uneven. While 18.6 percent (14.7 million people) lived in Istanbul province, 6.7 and 3.6 percent lived in Ankara and Izmir, respectively, in 2016. The population density of Istanbul, the densest province, is 2821 persons per square km, while it is only 12, in the least dense, Tunceli. All in all, while Istanbul is producing 30,5 percent of the national GDP, adding its immediate surrounding area increases the share to 41 percent in 2014.[1]

Regional imbalances in Turkey go back to the Ottoman Empire when the geographical location of Western Anatolia, especially the coastal areas like Izmir, Istanbul, and their hinterlands, gave these areas an essential role in the external trade of the country. Since then, trade and industry have always been more developed in these areas than in East Anatolia. With the foundation of the Republic of Turkey in 1923, the attention of the successive governments has shifted to Central Anatolia where the capital, Ankara, has been established. In order to reduce regional disparities, the Turkish authorities have implemented successive industrial plans to promote investments in public infrastructures and encourage private investments in the least favored parts of Anatolia.[2] After 1980, the influence of export base theory led to the decentralization of industrial activities from the metropolitan cities (Istanbul, Ankara, Izmir) and thus to the industrial expansion of the provinces adjacent to the metropolitan regions (Gezici and Hewings 2004). During the 1980s, Turkey adopted a regional policy strategy grounded on two main components: public investments in infrastructures and financial incentives (such as tax break, lump-sum payments) for the private sector to locate in the backward areas. Despite all the efforts, spatial inequalities remain a constant concern in Turkey today.

Turkish regional imbalances have been the subject of a vast amount of literature over the years. Research has addressed this problem in two ways, some work has utilized provin-

[1] We add Kırklareli, Tekirdağ, Kocaeli, Yalova, Sakarya and Bursa provinces.

[2] For a discussion on regional disparities in Turkey and government policies to tackle the issue see Celebioglu and Dall'erba (2010) and Gezici and Hewings (2004).

cial income data to study the income inequalities (Atalik, 1990; Gezici and Hewing, 2004), whereas most of the work has focused on provincial and regional convergence of income (Dogruel and Dogruel, 2003; Celbis and De Crombrugghe, 2016).

Although there have been a considerable number of studies dealing with productivity in the manufacturing sector (Krueger and Tuncer, 1982; Yildirim, 1989; Uygur, 1990; Aydogus, 1993; Gokcekus, 1997; Onder, Deliktas and Lenger, 2003, Atiyas and Bakis, 2014), there is shortage of regional level analysis. Karadag et al. (2005) study TFP change of the private and public sectors in the Turkish manufacturing industry in eighteen provinces from 1990-1998. Temel et al. (1999) use gross provincial product per worker for the period 1975-1990 and find evidence of polarization around specific highly industrialized regions. Most provinces tend to move toward a low productivity level, while a few moves toward a high productivity level, creating a divergence in productivity.

Coulibaly et al. 2007 is the closest work in spirit, to this paper. The authors assess the impact of urbanization on sectoral productivity from 1980-2000 by using manufacturing data and geographical, infrastructural and socio-economic data at province level. Their results suggest that localization (similar to specialization which measures how much local production is concentrated in a given activity) and urbanization economies[3], as well as market accessibility increase productivity. There has been criticism with this work, as their variables of interest are not instrumented thus suffer from endogeneity. Our work provides the first evidence from Turkey by using the best available methodological approaches given the data at hand.

3 Empirical Model

The economic geography literature emphasizes the importance of market size and market access in determining both, factor prices and the location of economic agents. Since Ciccone and Hall (1996), it is customary to measure the size of the local economy by density, which is the number of individuals per unit of surface area. Following the literature, we use the number of workers and estimate density as follows[4]:

[3] Urbanization is a composite measure of total number of firms within the province, urbanization rate, electricity consumption, the ratio for asphalt roads in villages. Market access has estimated the distance of province capital to nearest airport.

[4] Agglomeration mechanisms involve the size of the local economy which can be measured, depending on the mechanism, by employment, population, or production. Since these three variables are often highly

$$den_{p,t} = \frac{emp_{p,t}}{area_p} \tag{1}$$

To evaluate the relative influence of each of these variables, the logarithm of labor productivity is regressed on the logarithm of density:

$$p_{s,p,t} = \alpha + ln(den_{p,t}) + \gamma_s + \gamma_t + \epsilon_{s,p,t}. \tag{2}$$

where $p_{s,p,t}$ is labor productivity in sector s and province p at time t, $den_{p,t}$ is total employment density in province p, at time t, and error $\epsilon_{s,p,t}$ is an error term that captures local productivity shocks that are unexplained by the model.[5]

The panel structure of our data (81 provinces, 659 4-digit industries, and 9 years), allows us to introduce sector fixed-effects, γ_s which capture any sector-specific variables that affect all departments in the same way irrespective of time (e.g., labor productivity is on average higher in manufacture of machinery than in manufacture of brooms and brushes), and time fixed-effects,γ_t , picking up temporal variations affecting all provinces and all sectors equally (e.g. productivity gains from technological progress). Time fixed-effects also correct for the fact that all our variables are expressed in current Turkish Lira, and not deflated, which is less arbitrary than choosing a specific deflator (Combes and Gobillon, 2015).

The model can be estimated by simple Ordinary Least Squares (OLS) regressions if all the independent variables are observable and exogenous, but his hypothesis is rarely valid. Consequently, several estimation issues arise that we detail below.

3.1 Estimation Issues

Main issues when estimating agglomeration economies are unobserved heterogeneity and simultaneity. In the following sections, we propose methods to solve them.

correlated, identification of their effects separately is not possible. We use employment (instead of the population) as it better reflects the magnitude of local economic activity (Combes et Gobillon, 2015). We also use it to construct certain other local variables from employment only.

[5] Local population can also be used instead of density. However, as mentioned by Duranton (2015a) density-based measures of agglomeration are more robust to zoning idiosyncrasies.

3.1.1 Unobserved Heterogeneity

Some characteristics, unobserved by the econometrician, can be related to both labor productivity and some of the explanatory variables. In this case,$\epsilon_{s,p,t}$ is correlated with the independent variables; consequently, the OLS estimates of the coefficients are potentially biased, since the endogenous variables will partly capture the effect of unobserved characteristics. This issue is better known as the "unobserved heterogeneity" problem. In our specification, density is likely to be correlated with $\epsilon_{s,p,t}$.

To address the endogeneity arising from omitted variables, we add a number of variables drawn from the economic geography, which are standard in the literature. To distinguish density effects from pure scale effects, we use the province surface area, $area_p$. Density accounts for the market thickness, while land area measures its spatial extent. For instance, at a given density level, a larger area is likely to have more non-market interactions among agents than a smaller area because it is more populated.

The economic geography literature also suggests that proximity to large outlets induces greater profitability for firms. These markets may have a spatial scale larger than employment areas as argued by much of the recent literature (Fujita et al., 1999). It is customary to capture this market-access effect with a market potential variable à la Harris (1954). We differentiate market potential as foreign and domestic. Domestic market potential (DMP) for province p is defined as the sum of the other provinces' ($i \neq p$) density, divided by the road distance between provinces ($dist_{i,p}$):

$$DMP_{p,t} = \sum_{i \neq p} \frac{den_{p,t}}{dist_{i,p}}$$

The market potential of an area is defined with respect to all surrounding areas other than itself, first, to avoid multicollinearity issues and second, to identify separately the effects of the internal outlet (i.e., density) and external outlets (i.e., market potential). Foreign market potential (FMP) is estimated similarly. Instead of employment, we use GDP of countries to account for the size of their market and use the distance from province to country (via the closest trading port or border entry point).[6]

[6] Alternative DMP and FMP measures have been considered. For DMP, instead of density, the sum of salaries has been used to represent the size of the local market. For FMP, the sum of exports from the province p to the world has been used. Both alternative measures gave almost identical results.See Mayer and Head (2004), for more on the importance of domestic/foreign demand in agglomeration.

$$FMP_{p,t} = \sum \frac{GDP_{c,t}}{dist_{c,p}}$$

We also consider agglomeration economies arising from the sectoral distribution of economic activity. Local specialization is usually measured as the employment share of sector s in the economic activity of province p at date t,

$$spe_{s,p,t} = \frac{emp_{c,t}}{dist_{c,p}}$$

The diversity of the composition of local economic activity may also matter (Glaeser et al., 1992). We estimate the impact of local sectoral diversity via the inverse of a Herfindahl index, given by the sum of the squares of each sector's share in a given department:

$$div_{p,t} = [(\frac{emp_{s,p,t}}{emp_{p,t}})^2]^-1$$

New growth theories emphasize the role of human capital as a determinant of productivity (Lucas, 1988). In addition to its private benefits, having a higher amount of human capital (skill) can generate social benefits for urban workers. This effect can be tested by adding variables capturing the skills of the local labor force. We thus estimate now:

$$\begin{aligned} ln(p_{s,p,t}) = \alpha + \beta_1 ln(den_{p,t}) + \beta_2 ln(area_p) + \beta_3 ln(DMP_{p,t}) + \beta_4 ln(FMP_{p,t}) \\ +\beta_5 ln(spe_{s,p,t}) + \beta_6 ln(div_{p,t}) + \beta_7 HC_{p,t} + \gamma_t + \epsilon_{s,p,t} \end{aligned} \quad (3)$$

where $HC_{p,t}$ is a proxy for the share of skilled population in province p at date t. We define skilled population those with the university or higher degrees (ISCED4, ISCED5, and ISCED6). The coefficient is of a different nature to the other regression coefficients since it is not an elasticity. We test whether the introduction of human capital impacts labor productivity positively and whether it leaves the other estimated coefficients unchanged.

Local transportation infrastructure may also raise wages as they may lower exporting costs, make supplies cheaper or increase productivity. The effects of such "productive endowments" can we captured by the accounting as roads, airports, train lines, and others. However, such endowments are highly correlated with density and market potentials. Thus, adding such variables would increase the concerns about endogeneity. In order to avoid this problem, we follow Chauvin et al. (2016) and use climate-related amenities which are exogenous to the local economy and well measured.

3.1.2 Circular Causality

We follow the procedure proposed by Combes et al. (2008) to deal with circular causality and follow a two-step estimation.

In the first step, we estimate the impact of localization economies on productivity through the following equation:

$$lnp_{s,p,t} = \nu + \beta_1 lnspe_{s,p,t} + \gamma_{p,t} + \gamma_s + \epsilon_{s,p,t} \tag{4}$$

where $lnspe_{s,p,t}$, captures the effect of specialization in a given activity on productivity, $\gamma_{p,t}$ is a province-year fixed-effect which captures the influence of local non-sectoral variables on labor productivity. Year dimension of these fixed-effects correct for the fact that all our variables are expressed in current Turkish Liras, and not deflated. Also, differences in nominal wages between areas reflect differences in productivity and not the differences in “standard of living” which would be the case if real wages were used (Duranton, 2015).

γ_s is sector-fixed-effects capturing sector-specific variables that affect productivity across provinces. As we use province level aggregated data, we weight our results with the total number of workers of provinces.

In the second step, we regress the first-step predicted value $\gamma_{p,t}$ on local characteristics that can impact the productivity. To account for the local structure, we use density, surface area, market potential, diversity and year fixed effects, which can be instrumented separately from the first-step estimation:

$$\gamma_{p,t} = \alpha + \beta_1 ln(den_{p,t}) + \beta_2 ln(area_p) + \beta_3 ln(DMP_{p,t}) + \beta_4 ln(FMP_{p,t}) + \beta_6 ln(div_{p,t}) + \beta_6 HC_{p,t} + \gamma_t + \epsilon_{p,t} \quad (5)$$

This method is preferable for two reasons. First, we account separately for two distinct sector-province-year ($\epsilon_{s,p,t}$) and province-year ($\epsilon_{p,t}$) random terms. This allows us, in a second step, to tackle the endogeneity of density, market potential and human capital without addressing the endogeneity of the other variables, such as specialization.

Second, this procedure helps us separate externalities due to localization (first-step) from those due to urbanization (second-step) as well. This is particularly important for policy formulation, as it helps to determine whether policy focus should be on further developing existing sectors or encouraging the arrival of new activities to the region [7].

4 Data and Sample

In this paper, we use a novel administrative data set that was made public recently. We use social security data collected by the Social Security Institution (SGK), for administrative purposes. This monthly data set includes all of the workers affiliated with the social security system, both in private and public sector. Due to data privacy issues, the original individual-level data has been aggregated by the SGK at sector and province-level, to create this data set. Our sample includes monthly information on the number of workers, firms and daily wages, for 81 provinces, grouped according to Nace Rev.2 at 4-digit (659 sectors) for the period 2008-2016. We construct our sample by excluding Our sample includes 271,495 industry- province – year observations.

Since Ciccone and Hall (1996), it has become common practice to use long-lagged variables as instruments. Following the literature, we construct various instruments using Ottoman Empire population statistics of 1914 and the Turkish Republic's population census of 1927 and 1935. The last Ottoman census was conducted in 1905/1906. Population statistics of 1914 is an updated version of this census. The 1914 population data

[7] Using year dummies in the second step can increase the overall fit of the model and thus give a biased R^2. We check the robustness of our results by taking the average over the years for each variable at the national level and then subtracting each observation from the corresponding average. This detrending helps us eliminate the year effect, and thus makes it possible to drop year dummies. These results can be provided upon request.

used in this study were published for the first time by Karpat (1985), adapted to current administrative borders by Sakallı (2014).

Data on amenities (library, cinema, and more.), transportation statistics (road length, etc.), interprovincial distances come from Turkstat. International distance data come from CEPII, while GDP data come from the World Bank.

5 Results

In order to evaluate the agglomeration economies, we regress the logarithm of labor productivity, expressed as wages in our case, on the explanatory variables.

As we discussed in section 3.1.2, such a single-step estimation would require addressing the endogeneity of all variables simultaneously. Moreover, in single-step estimation, the variance of economic shocks has to be ignored when computing the covariance matrix of estimators. As shown by Moulton (1990), this creates large biases in the standard errors for the estimated coefficients of aggregate explanatory variables.

Instead, we proceed in two steps. First, we account separately for two distinct sector-province-year ($\epsilon_{s,p,t}$) and province-year ($\epsilon_{p,t}$) random terms. This allows us, in a second step, to tackle the endogeneity of density and other covariates without addressing the endogeneity of the other variables, such as specialization.

Our two-stage estimation consists of using province-year fixed effects estimated in equation (3) and using them as dependent variables in equation (4). The objective of this stage to assess the relative importance of endowments and between-industry interactions in explaining the province-year fixed effects. The province-fixed effects estimated in Eq. (3) are assumed to be a function of a year fixed effect, of local interactions between industries, and endowments.

5.1 Density

In order to evaluate relative influence of market size and market access, we begin by regressing the logarithm of labor productivity, expressed as wages, on the logarithm of employment per unit of surface area, denoted here, $\ln(den_{p,t})$ for province p at time t.

[Table 1 about here.]

The first column of Table 1 is the OLS estimation of the second stage, with density as the only explanatory variable for the province-year fixed dependent variable. When years are pooled together, density has an elasticity of 0.055. This means doubling the density produces $2^{0.055} - 1 = 3.9$ percent higher productivity all else equal. If the density of Iğdır (2.71, P25) were to increase to the density of Mersin (9.96, P75), its productivity would increase by 9.11 percent. When we run regressions separately for each year, we see that the coefficient moves around the coefficient we get when data are pooled, remaining on average consistent over the years.[8]

Here we ignore both reverse causality and individual unobserved heterogeneity to deal with spatial sorting. Our result echoes the elasticity of 0.06 found by Combes et al. (2008a) for a similar specification estimated on French employment areas over the period 1976–1998. For Spanish provinces, Martinez-Galarraga et al. (2008) estimate a sector-wide average elasticity of labor productivity to employment density which falls from 1860 to 1999, and which never exceeds 0.05. Ciccone and Hall (1996) obtain an estimate of 0.06 for American counties in 1988, and for the five largest EU-15 countries, Ciccone (2002) produces an estimate of approximately 0.05 for the end of the 1980s. Thus, elasticities seem to have similar orders of magnitude across studies.

5.2 Multivariate Analysis

In order to avoid endogeneity due to omitted variables, we add a number of controls to Table 1. Starting from the second column, we add controls that are standard in the agglomeration literature.

The elasticity of density ($\beta = 0.056$) increases to 0.064 as we add surface area (Column 2). The impact of surface area is significant, in line with other examples in the literature. When we add diversity, however, the inverse happens, and the elasticity drops to 0.043 (Column 3), while diversity remains insignificant.

Addition of domestic market potential and foreign market potential also have a similar impact, although foreign market potential seems to be much more powerful (Column 4-5).

8 Results are available upon request.

In order to account for human capital, we add the share of people with a university degree or higher. We do not include other levels as they remain insignificant once density is introduced. Note that the coefficient of human capital is different from other controls as it is a share and not an elasticity. In Column 6 we test whether the introduction of human capital impacts positively labor productivity, which seems to be the case.

In column 7 we bring together all the controls to test standard specification in the literature. When all controls are included, the coefficient of density drops to 0.054, which means that at a given surface area, doubling density increases productivity by 2.7 percent. While area remains statistically significant diversity is insignifcant. Once regressed simultaneously with domestic market potential, foreign market potential loses its significance. This common problem is due to high correlation between the two variables (≈ 0.93) and the lack of spatial variability of the foreign market demand by provinces (Redding and Venables, 2004; Combes et al., 2011).[9] That is why in the following sections, we work with domestic market potential. Lastly, the presence of human capital remains powerful.

5.3 Instrumental variables approach

As discussed earlier, the another important estimation issue is that some local characteristics are likely to be endogenous to local wages. For instance, employment areas receiving a positive technology shock may attract migrants. In such a context, as workers would be expected to move to productive cities, the quantity of labor would be endogenous. This leads to a positive correlation between the second-stage residuals and the density of employment. In this particular case, reverse-causality is going to bias the estimates upwards. Hence, endogeneity is a potentially serious concern for the second stage of the estimation (and all the more so since the direction of the bias is unclear). Regressors such as density, market potential or human capital are likely to be endogenous since they depend on workers' and firms' location decisions

In order to tackle this endogeneity issue, several instrumentation strategies have been proposed in the literature. Probably the most popular is using historical instruments that were proposed by Ciccone and Hall (1996), who argue that "contemporaneous" local shocks could cause endogeneity. This strategy rests on the hypothesis that historical values of population or density are relevant for today's levels as they are likely to hold

[9] See Redding and Venables (2004) or Combes et al. (2011) on the issue.

for very long periods, whereas they are exogenous to local outcomes of today (such as productivity, types of economic activities). In other words, local economic outcomes today are unlikely to be related to components of economic outcomes a long time ago that probably affected the historical population.

Following this strategy, we construct various instruments using Ottoman Empire population statistics of 1914 and the Turkish Republic's population census of 1927 and 1935. Our instruments are the logs of population density in 1914 and 1927, urban density growth (change in density between 1914-1927) and population potential (i.e., market potential in which population is substituted for employment) in 1935 and 1945.

As shown by the first-stage regressions presented in Appendix A, due to the strong inertia of the urban hierarchy in Turkey, population densities and market potentials at the beginning of the 21st century are still correlated to employment densities and market potentials in 1914 and 1927. However, it is unlikely that they are correlated to labor productivity in those years because Turkish economy has been subject to a wide range of productivity shocks triggered by war, significant population shifts between 1914-1924, and rural-urban migrations, which have profoundly affected the Turkish demography.

[Table 2 about here.]

In Table 2, we present IV results and along with statistics to test the validity of our instruments. Due to the historical character of our instruments, we are unable to instrument some of our observations. [10] In order to address concerns due to this change, we present the OLS estimation of the second stage, with density used as the only explanatory variable for a province-year fixed dependent variable with a reduced sample, which matches our instruments. When compared with the coefficient in Column 1 in Table 1, it can be seen that the coefficient of density does not change despite the decreases in the number of observations.

In the second column, we instrument the current density with lagged density in 1914 and the growth in urban density between 1914 and 1927. The elasticity decreases marginally to 0.054.

In Column 3, we add domestic market potential, and instrument both variables using

[10] We do not have population figures for three provinces (Kars, Ardahan, and Iğdır) as they were not under the Ottoman rule in 1914. These provinces joined the Turkish Republic in 1921.

the previous instruments and adding the log of domestic market potential calculated using population data in 1945.

When instrumented (Column 4) coefficient of density decreases to 0.052 while that of market potential increases to 0.054. This suggests that labor productivity is even more responsive to domestic demand ("home market effects") than to employment density. As noted by Combes et al. (2011), the more significant impact of market potential also captures agglomeration economies triggered by labor market pooling and knowledge spillovers diffusing over provincial boundaries.

Human capital is an essential determinant of productivity (Lucas, 1988). As the share of skilled labor is greater in denser areas (in particular cities), it is essential to distinguish the impact of these two factors on productivity. In order to do so, in Column 5 we include the share of the population with the university or higher degree $HC_{p,t}$ in province p in time t to capture the skill of the local labor force. The aim of introducing this variable is to see whether human capital impacts labor productivity positively and change the coefficient of density. Addition of human capital decreases the magnitude of density by almost 25

In Column 7 we add controls that were used previously and instrument them in Column 8. Once controls are added, the coefficient of density increases by some 11-30%. This result, although important, should be accepted with caution due to endogeneity concerns. In this final regression, we assume that controls are exogenous. Although area can be considered exogenous, the same cannot be said for the other two without making strong assumptions. Since instrumenting all these variables is impossible, endogeneity concerns would always remain. For this reason, we prefer coefficients in Column 6 where all of the variables are instrumented, taking care of endogeneity issues and potentially of omitted variable bias. Overall endogeneity leads to an overestimation of the density premium by 3-15 percent depending on the specification, which is lower than 20 percent found in Combes et al. (2011).

5.4 Infrastructure and Amenities

Local infrastructure and amenities can be factors that are correlated with city size and explain wages. In this section, we report three series of regressions testing their effect on wages. We also include geographical characteristics such as being located by the sea, having a large lake or high mountains in order to control for features that could have influenced location choices of the past populations (Combes and Gobillon, 2015).

First, we test for amenities. Roback (1982) highlighted the importance of consumption amenities for the willingness of consumers to pay for land and thus imply higher local land rents. As the local prices increase, firms use relatively less land which in turn decreases the marginal product of labor if the latter and land are not perfect substitutes.

[Table 3 about here.]

Column 2 of Table 3, we add three non-climatic amenity variables: number of cinema halls, hospitals and public libraries which are either amenities in themselves or proxies for amenities.[11] As predicted by spatial equilibrium involving utility equalization across locations, leisure amenities such as cinema halls and public libraries are associated with lower wages; they are on the other hand insignificant. Hospitals, however, are strongly significant and positive. Although theory would predict negative coefficient for such an amenity, this result might be due to reverse causation where more prosperous or more productive provinces can afford more hospitals.

Local transportation infrastructure may also matter. Recent literature shows that better provision of roads or better accessibility to the main road network of the country has a sizeable effect on both the volume and the value of exports of localities (Duranton, 2015b; Cosar and Demir, 2016). Whether the transport infrastructure also affects wages is, of course, another question. In Column 3, we add only the sum of rural roads and village roads per province.[12] First of all, estimated elasticity of wages with respect to density remains the same even when we add these variables. Secondly, infrastructure variables are not significant.

As shown in Cosar and Demir (2016) roads in municipality foster exports, which should have a positive effect on wages, transportation infrastructure may also lead to higher wages because they improve local productivity by facilitating local trades. On the other hand, better infrastructures may improve market access, lowering the price of imported goods, and thus the wages, especially if labor is mobile across places. In the Turkish

[11] Although not presented, we also tested for other amenities such as the number of theaters, museums, number of doctors and more. Results being insignificant, we chose to represent these variables for each amenity type.

[12] *Idem* for railways and motorways.

context, an increase in provincial road stock improves wages while the expansion of village road network is insignificant.

Finally, we add some controls for climate-related amenities following Chauvin et al. (2016) and account for average temperatures in January and its difference from the ideal temperature of 21.11 Celsius. We also account for the average yearly number of sunny days, and yearly rain volume.

In addition to our baseline regression, we add these climate-related amenities. Column 4 shows that a ten-degree Celsius difference between temperatures in winter from ideal temperature is associated with an increase of 0.5% in productivity. The number of days with sun or average rain, on the other hand, does not seem to matter.

Despite significant differences in climate between Turkish provinces, caused mainly by differences in elevation it does not seem to matter for productivity or prices. One explanation could be Turkey's economic divide which is west-to-east is already captured by other variables. Another explanation could be, that Turks are not wealthy enough to be willing to pay a significant premium to live in places with more moderate climate conditions. Although evidence from developed countries show that climate is an important amenity (Cheshire and Magrini, 2006; Rappoport, 2007), this seems to not matter for developing countries as it has been shown for Colombia (Duranton, 2015), China or India (Chauvin et al. 2016).

6 Conclusion

In this paper, we adopt a unified framework to examine the determinants of productivity differences across Turkish provinces. When correctly identified the elasticity of productivity with respect to density is about 0.038-0.054 percent which is higher than in developed countries and around the levels observed in developing countries. Economies of density play an important role in explaining differences in local wages. We find that domestic market potential and human capital externalities matter.

One important issue to note is that, due to data availability, we used aggregate data as opposed to individual-level data used. The absence of individual data in developing countries makes it impossible to implement state of the art methodology that takes into account the sorting of workers across spaces. Although inclusion of the education variable takes care of skill composition of the workforce, sorting across space based on unobservable heterogeneity is not entirely taken care of. This means that potentially the coefficient of density is overestimated.

We find that the elasticity of productivity with respect to density in Turkey is higher than developed countries and around the levels observed in developing countries. If however, the coefficient is overestimated due to data constraints, then it would mean that agglomeration externalities in Turkey are weaker compared to its urbanization level and economic development. In other words, the advantages of agglomeration are limited by disadvantages and costs such as traffic, pollution and more. There are a couple of potential explanations as to why it may be the case.

The first potential factor is the bad organization of urbanization in Turkey and current city size distribution not optimal. Over-population and unorganized urbanization might increase the costs of agglomeration which, in their turn might decrease the coefficient of density.

A second potential factor might be the problems concerning the land market. Land markets in developing cities are characterized by a duality between land used with appropriate property titles and leases and squatted land. Following a conjecture by De Soto (2000), recent empirical research has focused on the effects of the lack of effective, formal property titles that could prevent residents of squatter settlements from using their house as collateral. Informal land markets may thus be a significant barrier to enterprise development.

A third potential factor might be the effects of the favoritism by governments towards the largest cities. Although the reasons for primate city favoritism are debated (Ades and Glaeser 1995; Henderson 2005), there is little doubt that such favoritism occurs in many different ways. As argued in Duranton (2008), primate city favoritism harms the favored primate city by making it more significant than it should be. It also harms smaller cities, which are, in effect, heavily taxed. The gap that is created between the primate city and other cities may also have adverse dynamic effects because, for most educated workers, there is nowhere to go except to stay in this primate city. As a result, this may reduce the circulation of knowledge across cities. Reducing primate city favoritism and providing smaller cities with better local public goods (including education and health) is undoubtedly a part of any solution.

A fourth potential factor might be the strong rural-urban migration flow in Turkey. Massive influx from the rural zones to urban areas and especially towards bigger cities might have dampened the benefits of density, as these migrants lacked skills that could be put into use in productive activities mostly exercised in urban areas. The arrival of the rural population with very limited skill-sets increased the supply of low-skilled labor far faster than the increase in demand for such labor, bringing down the average skill-level in urban areas. In Hecker-Ohlin perspective, over-abundance of low-skilled labor may have even increased the preference for labor-intensive activities, which could also have decreased the gains of increased concentration of production factors.

7 References

Ades, A. F. and E. L. Glaeser. 1995. “Trade and Circuses: Explaining Urban Giants.” The Quarterly Journal of Economics 110(1):195–227.

Albouy, D., F. Leibovici, and C. Warman. 2013. “Quality of Life, Firm Productivity, and the Value of Amenities across Canadian Cities.” Canadian Journal of.

Atalik, G. 1990. “Some Effects of Regional Differentiation on Integration in the European Community.” Papers of the Regional Science Association.

Atiyas, İ. and O. Bakış. 2014. “Aggregate and Sectoral TFP Growth in Turkey: A Growth Accounting Exercise (Türkiye’de Toplam ve Sektörel Toplam Faktör Verimliliği Büyüme Hızları: Bir Büyüme.” İktisat, İşletme ve Finans.

Celbis, MG and D. Crombrugghe. 2016. “Internet Infrastructure and Regional Convergence: Evidence from Turkey.” Papers in Regional Science.

Celebioglu, F. and S. Dall’erba. 2010. “Spatial Disparities across the Regions of Turkey: An Exploratory Spatial Data Analysis.” The Annals of Regional Science.

Chauvin, Juan Pablo, Edward Glaeser, Yueran Ma, and Kristina Tobio. 2016. “What Is Different about Urbanization in Rich and Poor Countries? Cities in Brazil, China, India, and the United States.” Journal of Urban Economics 0:1–61.

Cheshire, PC and S. Magrini. 2006. “Population Growth in European Cities: Weather Matters–but Only Nationally.” Regional studies.

Ciccone, A. 2002. “Agglomeration Effects in Europe.” European Economic Review.

Ciccone, Antonio and Robert E. Hall. 1996. “Productivity and the Density of Economic Activity.” American Economic Review 86(1):54–70.

Combes, Pierre Philippe, Sylvie Demurger, and Shi Li. 2015. “Migration Externalities in Chinese Cities.” European Economic Review 76:152–67.

Combes, Pierre Philippe, Gilles Duranton, and Laurent Gobillon. 2008. “Spatial Wage Disparities: Sorting Matters!” Journal of Urban Economics 63(2):723–42.

Combes, PP, G. Duranton, and L. Gobillon. 2008. “Spatial Wage Disparities: Sorting Matters!” Journal of Urban Economics.

Combes, PP and L. Gobillon. 2014. “The Empirics of Agglomeration Economies.”

Combes, PP, M. Lafourcade, and JF Thisse. 2011. “The Rise and Fall of Spatial Inequalities in France: A Long-Run Perspective.” Explorations in Economic.

Coşar, AK and B. Demir. 2016. “Domestic Road Infrastructure and International Trade: Evidence from Turkey.” Journal of Development Economics.

Coulibaly, S., U. Deichmann, and SV Lall. 2007. “Urbanization and Productivity: Evidence from Turkish Provinces over the Period 1980-2000.”

Duranton, G. 2015. “Growing through Cities in Developing Countries.” The World Bank Research Observer. Duranton, Gilles. 2016. “Agglomeration Effects in Colombia.” Journal of Regional Science 56(2):210–38.

Gezici, F. and GJD Hewings. 2004. “Regional Convergence and the Economic Performance of Peripheral Areas in Turkey.” Review of Urban and Regional.

Henderson, JV. 2003. “Urbanization, Economic Geography and Growth.” Handbook of Economic Geography. Henderson, JV. 2005. “Urbanization and Growth.” Handbook of economic growth.

Karadağ, M., Ö. Önder, and E. Deliktaş. 2005. “Growth of Factor Productivity in the Turkish Manufacturing Industry at Provincial Level.” Regional Studies.

Krueger, AO and B. Tuncer. 1982. “Growth of Factor Productivity in Turkish Manufacturing Industries.” Journal of Development Economics.

Krugman, Paul. 1991. “Increasing Returns and Economic Geography.” The Journal of

Politial Economy 99(3):483–99.

Lucas, RE. 1988. “On the Mechanics of Economic Development.” Journal of monetary economics.

Martínez-Galarraga, J., E. Paluzie, and J. Pons. 2008. “Agglomeration and Labour Productivity in Spain over the Long Term.” Cliometrica.

Michaels, Guy, Ferdinand Rauch, and Stephen J. Redding. 2012. “Urbanization and Structural Transformation.” Quarterly Journal of Economics 127(2):535–86.

Moretti, E. 2004. “Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data.” Journal of econometrics.

Önder, AÖ and ERUL DELIKTAS. 2003. “Efficiency in the Manufacturing Industry of Selected Provinces in Turkey: A Stochastic Frontier Analysis.” Emerging Markets Finance.

Rappaport, J. 2007. “Moving to Nice Weather.” Regional Science and Urban Economics.

Rauch, JE. 1993. “Does History Matter Only When It Matters Little? The Case of City-Industry Location.” The Quarterly Journal of Economics.

Redding, S. and AJ Venables. 2004. “Economic Geography and International Inequality.” Journal of international Economics.

Roback, J. 1982. “Wages, Rents, and the Quality of Life.” Journal of political economy.

Temel, T., A. Tansel, and PJ Albersen. 1999. “Convergence and Spatial Patterns in Labor Productivity: Nonparametric Estimations for Turkey.”

Turkish Statistical Institute, www.turkstat.gov.tr, access: 17.02.2017

http://data.worldbank.org/indicator/NY.GDP.PCAP.CD, access: 17.02.2017

Table 1: Productivity and density: Multivariate analysis

	1	2	3	4	5	6	7
LnDen	0.055***	0.064***	0.043***	0.054***	0.043***	0.040***	0.054***
	(0.003)	(0.003)	(0.005)	(0.002)	(0.005)	(0.002)	(0.006)
LnArea		0.057***					0.031***
		(0.017)					(0.010)
LnDiv			0.082***				-0.012
			(0.026)				(0.023)
LnDMP				0.050***			0.058**
				(0.019)			(0.022)
LnFMP					0.195***		-0.043
					(0.064)		(0.112)
HC						7.495***	4.898***
						(0.739)	(0.727)
N	726	726	726	726	726	726	726
R^2	0.806	0.874	0.827	0.831	0.823	0.910	0.940

Notes: (i) Ordinary Least-Squares. (ii) Standard errors in brackets,robust to departmental clusters. (iii) Significance levels: * p<0.1, ** p<0.05, *** p<0.01. (iv) Dependent variable is province-year fixed effect estimated in Equation 3.

Table 2: Productivity and density: IV estimates

	1	2	3	4	5	6	7	8
	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)
LnDen	0.055***	0.054***	0.054***	0.052***	0.040***	0.038***	0.052***	0.043***
	(0.003)	(0.005)	(0.002)	(0.004)	(0.001)	(0.002)	(0.005)	(0.010)
LnDMP			0.051**	0.054**	0.049***	0.039***	0.051***	0.031***
			(0.019)	(0.026)	(0.014)	(0.014)	(0.014)	(0.014)
HC					7.112***	6.718***	4.770***	3.786**
					(0.470)	(1.053)	(0.692)	(1.724)
LnArea							0.032***	0.026*
							(0.009)	(0.015)
LnDiv							-0.016	0.040
							(0.023)	(0.034)
N	726	726	726	726	594	594	594	594
R^2	0.806	0.806	0.831	0.830	0.982	0.928	0.941	0.931
Cragg-Donald Stat		889.3		452.3		118.5		52.33
P-value Hansen test		0.340		0.367		0.285		0.953
Shea's Partial(LnDen)		0.965		0.915		0.823		0.682
Shea's Partial(LnDMP)				0.778		0.795		0.514
Shea's Partial(ISCED56)						0.557		0.367

Notes: (i) IV 2SLS. Density instrumented in column (IV1): excluded instruments are the log of population density in 1914, log of density growth between 1914-1945; Density and domestic market potential instrumented in column (IV2): excluded instruments are the log of population density in 1914, log of density growth between 1914-1945 and log of domestic market potential in 1945; Density, domestic market potential and ISCED56 instrumented in column (IV3): excluded instruments are the log of population density in 1914, log of density growth between 1914-1945,log of domestic market potential in 1945 and ratio of students to population in 1927; first-stage regressions are reported in the Appendix A. (ii) Standard errors in brackets,robust to departmental clusters. (iii) Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$. (iv) Dependent variable is province-year fixed effect estimated in Equation 3.

Table 3: Infrastructure and Amenities

	1	2	3	4	5
LnDen	0.040***	0.023***	0.042***	0.056***	0.048***
	(0.001)	(0.009)	(0.007)	(0.007)	(0.008)
LnDMP	0.049***	0.058***	0.037***	0.047***	0.040***
	(0.014)	(0.016)	(0.008)	(0.014)	(0.014)
HC	7.112***	6.428***	5.753***	5.611***	4.509***
	(0.470)	(0.924)	(0.855)	(0.850)	(0.968)
Hospitals		0.034***			0.039
		(0.020)			(0.025)
Public Libraries		-0.022			-0.048***
		(0.014)			(0.013)
Cinemahalls		0.004			-0.004
		(0.010)			(0.010)
Provincial Roads			0.058***		0.065**
			(0.018)		(0.026)
Village Roads			-0.030		-0.018
			(0.019)		(0.025)
January				0.007***	0.005***
				(0.002)	(0.002)
Sunny Days				0.014	0.012
				(0.010)	(0.010)
Average Rain				0.000	0.001**
				(0.000)	(0.000)
Geography	Y	Y	Y	Y	Y
N	594	594	594	594	594
R^2	0.982	0.940	0.881	0.874	0.780

Notes: (i) Ordinary Least Squares (ii) Standard errors in brackets, robust to departmental clusters. (iii) Significance levels: * p<0.1, ** p<0.05, *** p<0.01. (iv) Dependent variable is province-year fixed effect estimated in Equation 3. (v) Number of hospitals, number of public libraries, number of cinema halls are computed as 1+ number in 2007 to accommodate the log transformation. Provincial and village roads are the sum of kilometers in 2007 in logs. Number of sunny days and annual average of rain (kg/m2) of 1950-2015 are in logs. January temperatures are average of 1950-2015 and are in levels. (vi) Geographic controls include a dummy for access to the coast, presence of a large lake and mountain.

Table 4: First Stage - Column 2

	LnDen
LnDen(1914)	1.066***
	(0.251)
DensityGrowth(1914-1945)	0.00379***
	(0.00104)
N	726

Notes: (i) Ordinary Least-Squares. (ii) Standard errors in brackets,robust to departmental clusters. (iii) Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table 5: First Stage - Column 4

	LnDen	LnDMP
LnDen(1914)	1.052***	0.0159
	(0.223)	(0.0234)
DensityGrowth(1914-1945)	0.00443***	-0.000172
	(0.00114)	(0.000218)
LnDMP(1945)	-1.500	1.592***
	(1.028)	(0.144)
N	594	594

Notes: (i) Ordinary Least-Squares. (ii) Standard errors in brackets,robust to departmental clusters. (iii) Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table 6: First Stage - Column 6

	LnDen	LnDMP	ISCED56
LnDen(1914)	0.989***	0.00312	0.0002
	(0.216)	(0.0171)	(0.00134)
DensityGrowth(1914-1945)	0.00528***	0.00005	0.00002***
	(0.00124)	(0.0002)	(0.00007)
LnDMP(1945)	-2.195**	1.449***	0.00166
	(0.950)	(0.130)	0.00166
Ratio School(1927)	23.17*	5.631***	-0.0816
	(12.67)	(1.200)	(0.0679)
N	594	594	594

Notes: (i) Ordinary Least-Squares. (ii) Standard errors in brackets,robust to departmental clusters. (iii) Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table 7: First Stage - Column 8

	LnDen	LnDMP	ISCED56
LnDen(1914)	0.684***	-0.0113	0.000553
	(0.197)	(0.0307)	(0.00103)
DensityGrowth(1914-1945)	0.00351***	-0.000100	0.00002***
	(0.00107)	(0.000192)	(0.00007)
LnDMP(1945)	-2.731***	1.348***	-0.00329
	(0.644)	(0.146)	(0.00363)
Ratio School(1927)	20.86***	6.471***	-0.00358
	(7.462)	(1.635)	(0.0378)
LnArea	-0.0151	0.0482	0.00384**
	(0.294)	(0.0544)	(0.00157)
LnDiv	2.128***	0.200*	0.00402*
	(0.426)	(0.104)	(0.00212)
N	594	594	594

Notes: (i) Ordinary Least-Squares. (ii) Standard errors in brackets,robust to departmental clusters. (iii) Significance levels: * p<0.1, ** p<0.05, *** p<0.01.