



HAL
open science

Agglomeration Effects in a Developing Economy: Evidence from Turkey

Cem Özgüzel

► **To cite this version:**

Cem Özgüzel. Agglomeration Effects in a Developing Economy: Evidence from Turkey. 2020. halshs-02878368

HAL Id: halshs-02878368

<https://shs.hal.science/halshs-02878368>

Preprint submitted on 23 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



PARIS SCHOOL OF ECONOMICS
ECOLE D'ECONOMIE DE PARIS

WORKING PAPER N° 2020 – 41

Agglomeration Effects in a Developing Economy: Evidence from Turkey

Cem Özgüzel

JEL Codes: R12, R23, J31.

Keywords: local labor markets, spatial wage disparities, developing country,
Turkey



Funded by a French government subsidy managed by the
ANR under the framework of the investissements d'avenir
programme reference ANR-17-EURE-001

Agglomeration Effects in a Developing Economy:

*Evidence from Turkey**

Cem Özgüzel †

OECD

Paris School of Economics

Latest Update: June 11, 2020

Keywords: local labor markets, spatial wage disparities, developing country, Turkey

JEL Classification Numbers: R12, R23, J31

*I thank Güney Celbis, Pierre-Philippe Combes, Hadi Esfahani, Davide Luca, Sandra Poncet, Ariell Reshef and Gregory Verdugo, seminar participants at University of Paris 1, University of Strasbourg and at the 25th Economic Research Forum (ERF) Conference (Kuwait) for useful feedback. This paper received the Best Paper Award at the 25th ERF Conference (2019). I want to thank the Social Security Institution (SGK) and Turkish Statistical Institute (Turkstat) for kindly providing the data. I acknowledge financial support from the ERF. All errors are mine.

†Email: cem.ozguzel@oecd.org

Abstract

Productivity differences across Turkish provinces is one of the highest among the OECD countries. In this paper, I estimate agglomeration effects for Turkish provinces to shed light on the causes of productivity differences and provide evidence on the importance of such effects in a developing country context which literature needs. I use a novel administrative dataset recently made available at NUTS-3 level, for 81 provinces of Turkey for the period 2008-2013 and carry out a two-step estimation. Using a variety of panel data techniques and historical instruments to deal with estimation concerns, I estimate an elasticity of labor productivity with respect to the density of 0.057-0.06, which is higher than in developed countries and around the levels observed in developing countries. I find that domestic market potential matters even more than density and is the most significant determinant of the productivity differences across Turkish provinces. Finally, in stark contrast with the evidence coming from developed countries, I do not find any effects for positive sorting of workers across provinces. This finding suggests that urbanisation patterns may be operating differently in developing countries, indicating the need for further evidence from such countries.

1 Introduction

Despite an extensive literature on the productivity gains associated with large cities in developed economies, little is known about this topic in the developing countries. Investigating agglomeration economies in the rest of the world is necessary for a number of reasons. First, the growth rate of the world's urban population is being driven by urbanization outside the developed world (OECD, 2020). Given the importance of cities in national economies and the millions of people who live in them, understanding the drivers of agglomeration economies are policy-relevant for boosting national productivity, reducing regional disparities and improving the well-being of people. Second, one can not assume existing developed world models are applicable to agglomeration economies of the developing world (Chauvin et al., 2017). For instance, the rapid urbanization observed in developing countries in the second half of the past century stands in stark contrast to the stability seen in the developed world (Glaeser and Henderson, 2017). The rapid growth in urbanization rates, the differences in the growth of cities, institutional or infrastructural qualities may affect the size and the extent of the gains associated with larger and denser cities. In light of these factors, this paper aims to further our understanding of these critical differences by providing evidence from a Turkey, a highly urbanized developing country.

Spatial differences in productivity are driven through three main channels: i) skill composition of workers, ii) local non-human factors, iii) interactions between workers or firms. First, differences in productivity across locations (i.e. cities or regions) could be due to differences in the skill composition of the workers. Workers with higher skills may sort into denser areas if, for instance, skill-intensive industries are not evenly distributed across areas, or if they value cultural amenities more than low-skilled workers. Second, the differences in productivity could be due to local non-human factors. For instance, workers in some locations may have a higher marginal product because of geographical features such as a favourable location (like a port or a bridge on a river), a climate more suited to economic activity, or access to natural resources. Third, local interactions between workers or firms can lead to productivity gains, which is known as Marshallian externalities (see Duranton and Puga (2004) for a review). Thus, to understand spatial differences in productivity, all three channels have to be quantified and considered together.

This paper takes up the challenge of considering all of these factors simultaneously to study the determinants of wage disparities, a key measure of productivity¹, across Turkish provinces.² Since 1950, Turkey has experienced urbanization much faster than developed countries due to massive rural-urban migration. Today, 75% of the Turkish population lives in cities, making it a highly urbanized country

¹Wages are usually proportional to labor productivity. By using the wages as a measure to compare relative productivity differences within a country, the literature assumes that the proportion does not vary across regions within the country (Combes and Gobillon, 2015).

²Provinces (*il* in Turkish), correspond to the NUTS-3 (Nomenclature of Territorial Units for Statistics) level in the Eurostat classification of regions.

(World Bank).³ In terms of GDP per capita, Turkey has the highest regional disparity among the OECD countries (OECD, 2018), with substantial spatial inequalities existing across almost every metric (e.g., production or life expectancy). The rapid urbanization experience and significant spatial differences in productivity make Turkey an excellent context for understanding the sources and consequences of agglomeration economies in a highly urbanized developing country.

The analysis uses a novel administrative dataset based on social security records which covers almost the entire universe of workers across all industries, working in both the public and private sector. In addition to its extensive coverage and reliability, the dataset offers the advantage of measuring agglomeration effects at the province level, a finer geographical unit than any other available data in Turkey. Due to data privacy concerns, the individual observations are aggregated at the industry-province-year level, which makes it impossible to account for spatial sorting of workers based on differences in their observable and unobservable characteristics. I circumvent this problem by complementing my analysis using the Household Labor Force Survey, which provides individual-level data yet at the regional level. Finally, to address endogeneity bias and reverse causality, I use instruments based on past settlement patterns using historical data from the last census from Ottoman Empire and early census of the Turkish Republic.

First, when controlling for local characteristics and addressing reverse causality, I find an elasticity of wages with respect to employment density to be around 0.056-0.06. This means that doubling the employment density in a province increases the average wages (and productivity) of workers by 3.8-4.2%. Compared to other developing countries, this elasticity is lower than those estimated for China (Combes et al., 2015; Chauvin et al., 2017) and India (Chauvin et al., 2017), around those estimated for Brazil (Chauvin et al., 2017) and Colombia (Duranton, 2016), and higher than Ecuador (Matano et al., 2020). Given the levels of urbanization in these countries and Turkey, this elasticity fits precisely where it would be expected, hinting decreasing benefits of agglomeration economies for countries with higher levels of urbanization rates.

Second, I find a positive and strong effect of domestic market potential, on labour productivity. The estimated coefficient is around 0.091-0.1, which is double that of density, suggesting that having access to other markets is the most significant determinant of the productivity differences across Turkish provinces. The estimated elasticity implies that if the market potential of a province doubles, the wages increase by 6.5%. This result corroborates findings from other developing countries which show that productivity gains associated with larger market potential are stronger than developed countries and that they matter for explaining spatial productivity differences (Combes and Gobillon, 2015).

³Turkey went through rapid urbanization starting from the 1950s due to large rural-urban migration. While 34% of the population lived in cities in 1950, this number increased to 42% in 1975, 53% in 1985, 64.9% in 2000 and it is at 74.5% in 2017 (World Bank). Most of the internal migrants were low-skilled agricultural workers (Kirdar and Saracoglu, 2008). While 62.5% of the Turkish labor force was employed in agriculture in 1980, only 18% remained in this sector as of 2018 (Turkstat).

Third, I do not find any effects of sorting of workers across locations according to their observable skills. This result is in sharp contrast with what is usually observed for developed countries, where a large fraction of the explanatory power of city effects arises from the sorting of workers (Combes and Gobillon, 2015). It is, however, very much in line with the results for China (Combes et al., 2015, 2020), which suggests that urbanization patterns may be operating differently in developing countries, supporting the need for further evidence from such countries (Chauvin et al., 2017).

Finally, I find a weak relationship between productivity (wages) and amenities. This result echoes the findings in other developing countries and can be explained by either the high correlation between density and amenities (Duranton, 2016); or that workers in developing countries are not rich enough to forgo part of their income to live in areas with better amenities (Chauvin et al., 2017).

This paper makes several important contributions. First, the paper is the first to estimate agglomeration economies in Turkey and consider a broad set of factors that influence productivity differences simultaneously in a developing country context. Considering a broad set of explanatory factors allows quantifying the magnitudes associated with each explanation and comparing their relative importance. Given the importance of productivity as a driver of growth, understanding which local factors make a given worker more productive is crucial for formulating policies to reduce regional disparities. Second, estimated magnitudes for local characteristics such as density or domestic market potential indicate stronger effects compared to those observed in developed countries. These results further our understanding of the importance of local characteristics given the urbanization rate of a country. Third, the absence of positive sorting effects and a weak relationship between wages and amenities highlights a fundamental difference between developed and developing countries regarding the drivers of urbanization. These findings corroborate recent evidence from developing countries that show that while the main mechanisms of urban economies are present in the developing world, the current models need to be extended to capture the differences observed in the developing part of the world which has experienced fast urbanization rates.

The rest of the paper is organized as follows. Next section provides details about the Turkish context (Section 2), followed by Section 3 which presents the data used in the analysis. Section 4 presents the empirical strategy and discuss the identification issues. Section 5 provides estimates on density and other determinants of productivity. Section 6 concludes the paper.

2 The Turkish context

Turkey has a population of about 81 million people over an area of 783 thousand square kilometres (Turkstat, 2019). With a Gross Domestic Product (GDP) per capita of USD 10546 in 2017, according to the World Bank Turkey is an upper-middle-income developing country. However, this wealth is not

equally distributed across its regions (See Figure 1). While the GDP per capita in Istanbul was USD 17827, it was only USD 3489 in Ağrı (Turkstat, 2019).⁴ These differences are multiplied even further at the aggregate level since population distribution is also uneven. In 2017, while 18.6% (15.1 million people) lived in Istanbul province, 6.7% (5.4 million) and 5.2% (4.3 million) lived in Ankara and Izmir, respectively. The population density of Istanbul, the densest province, is 2892 persons per square km, while it is only 11, in the least dense, Tunceli (Turkstat, 2019). All in all, while Istanbul is producing 31.5% of the national GDP, adding its immediate surrounding area increases the share to 39% in 2017 which creates large imbalances across regions in terms of production and income (Turkstat, 2019).⁵

[Figure 1 about here.]

2.1 Literature on regional differences in Turkey

Turkish regional imbalances has been the focus of a number of papers over the years. While some papers used provincial income data to study the spatial income inequalities (Atalik, 1990; Gezici and Hewings, 2004; Luca, 2016), others focused on provincial and regional convergence of income (Kirdar and Saracoglu, 2008; Celbis and de Crombrugghe, 2018).

Although there has been considerable number of studies dealing with the determinants of overall productivity at the national level (Altug et al., 2008; Ismihan and Metin-Ozcan, 2009; Atesagaoglu et al., 2017) or in the manufacturing sector (Krueger and Tuncer, 1982; Onder et al., 2003; Atiyas and Bakis, 2014), regional level analysis has been limited. Few examples include Metin et al. (2005) who study TFP growth in the Turkish manufacturing industry in eighteen provinces from 1990 to 1998 or Temel et al. (1999) who use gross provincial product per worker for the period 1975-1990. These papers find evidence of concentration of high productivity activities in a few highly industrialized regions while most provinces tended to move towards low productivity activities, creating regional divergence in terms of productivity.

Determinants of spatial productivity in Turkey has not been addressed in the literature. Coulibaly et al. (2007) is the closest work in spirit, to this paper. The authors assess the impact of urbanization on sectoral productivity between 1980 and 2000 by using manufacturing data and geographical, infrastructural and socio-economic data at the province level. Their results suggest that localization (similar to specialisation which measures how much local production is concentrated in a given activity) and urbanization economies, as well as market accessibility increase productivity. Authors do not, however, deal with identification issues such as endogeneity of local determinants of agglomeration economies or the sorting of workers.

This paper is the first causal evidence from Turkey focusing on the spatial productivity differences

⁴Regional imbalances in Turkey go back to late Ottoman Empire. For more details on the historical background of geographical disparities see Appendix Section A.

⁵The surrounding area of Istanbul includes Kırklareli, Tekirdağ, Kocaeli, Yalova, Sakarya and Bursa provinces.

through agglomeration literature perspective. The novel data that I use in this paper (see the next section), also makes it the first paper that analysis productivity differences at province level covering the period after 2000.

3 Data and Sample

The paper uses different data sources to estimate the determinants of spatial differences in wages. This section presents the main data used in the analysis.

3.1 Data

3.1.1 Social Security Data

In this paper, I use a novel administrative data set that was made available to researchers recently. This data set is collected by the Social Security Institution (*Türkiye Cumhuriyeti Sosyal Güvenlik Kurumu* - SGK) and are based on administrative records for all the workers affiliated to the social security system. It covers employment in all of the industries, in the private and public sector.⁶

Due to data privacy issues, the raw individual-level data is aggregated by the SGK by sector, province and year. Thus the data includes yearly information on the number of workers, the total number of days worked, the number of firms and total payments received (wages and benefits) by the workers, for 81 provinces, grouped according to Nace Revision 2 at 4-digit sector level (659 sectors) from 2008 to 2013. The data is further disaggregated by job contract-type (temporary vs permanent), by sex (male vs female) and legal status (public vs private-owned).

The novelty of this data is that it makes analysis at province-level possible. The literature focusing on individual or firm-level outcomes has been limited to analysis at region level (NUTS-2) due to data availability.⁷ This dataset makes it possible, for the first time for the period since 2000, to focus on productivity differences at a geographically more disaggregated level. This is also the first paper to use this dataset.

The final sample for analysis includes male and female workers employed in private sector⁸, holding temporary and permanent job contracts in Turkey from 2008 through 2013. This amounts to 168 904 industry-province-year observations, which I use to estimate the province-year fixed-effects in the first step.

⁶The data covers all employment with compulsory insurance in the private and public sector under Article 4-1/a and 4-1/c of Act 551. For the year 2013, this corresponds to 13,1 million formal employees in private (12,5 million) and in public (650 thousand) sector. It does not include apprentices (321 thousand), those who work abroad but are affiliated with the Turkish Social Security System (35 thousand), the agricultural sector (64 thousand) and voluntary based insured partially employment (230 thousand). It does not include the self-employed (2.9 million) who are covered under Article 4-1/b of Act 5510.

⁷Two main data sources used in studies focusing on the Turkish labor market are Household Labor Force Survey (*Hanehalkı İşgücü Anketi*) and Annual Industry and Service Statistics (*Yıllık Sanayi ve Hizmet İstatistikleri*) which are provided by Turkish Statistical Institute (*Türkiye İstatistik Kurumu*). Both datasets allow identification at NUTS-2 level.

⁸I exclude workers in public sector as their wage is not set by market mechanism.

3.1.2 Household Labor Force Survey

I complement the main analysis by using individual level data obtained from the Household Labor Force Survey (*Hanehalkı İşgücü Anketi*, in Turkish, LFS henceforth) prepared by the Turkish Statistical Institute (*Türkiye İstatistik Kurumu*, in Turkish, Turkstat henceforth). The LFS is representative of the total population in Turkey and is used to follow the state of the economy and the labor market, both at the regional and national level. The LFS includes annual information on economic activity, occupation, employment status, hours worked (for persons employed in formal or informal sector), unemployment, education and more.

In the period of my analysis, each survey wave included around 135 thousand households, covering 500 thousand individuals. Its high level of detail and large sampling size makes it an important data source for research on the regional and national labor markets in Turkey (e.g., Tumen, 2016; Balkan and Tumen, 2016; Baslevant and Onaran, 2003, 2004).

The main shortcoming of this dataset is that it allows analysis only at regional level (NUTS-2). In my analysis, I prefer using the social security data, which will enable me to study the local interactions at a lower geographical unit (i.e., province-level, NUTS-3) and use the LFS as a complement to address identification issues.

3.1.3 Historical Data

Since Ciccone and Hall (1996), it is standard practice to use long-lagged variables as instruments for local characteristics. Following the literature, I construct various instruments using Ottoman Empire population statistics of 1914 and the Turkish Republic's population censuses of 1927, 1935 and 1945. The last Ottoman census was conducted in 1905/1906 and population statistics of 1914 is an updated version of this census. The 1914 population data used in this study were published for the first time by Karpat (1985), adapted to current administrative borders by Sakalli (2019).⁹ I complement this data by digitizing published census reports for the period 1927, 1935 and 1945 which come from Turkstat. In addition to the population statistics, these data also include information on occupations, number of students, number of schools and much more.

I use these data to calculate the past population densities and past domestic market potential. I also use the number of enrolled male students to elementary schools and high schools in 1927. Finally I compute the foreign-market potential using historical GDP data for Turkey's trading partners coming from the Maddison Project Database (Bolt and van Zanden, 2014).

⁹Between 1914 and 2007, the number of districts, names, and their borders has changed considerably. I use the correspondence between past and current administrative boundaries prepared by Sakalli (2019).

4 Empirical Strategy

This section presents the framework used for estimating the agglomeration effects in Turkey. It also discusses possible identification issues arising from reverse causality and selection bias due to sorting of workers by their ability.

4.1 Econometric Equation

To evaluate the impact of agglomeration economies on productivity, I use a two-step estimation strategy that is standard in the literature:

$$\log w_{pst} = \alpha + \delta \log Spe_{pst} + \gamma_s + \gamma_{pt} + \varepsilon_{pst} \quad (1)$$

$$\gamma_{pt} = \nu + \beta_1 \log Den_{pt} + \theta X_{pt} + \phi Z_p + \gamma_t + \varepsilon_{pt} \quad (2)$$

In the first step (Equation 1), I regress the log average daily wages ($\log w_{pst}$) in province p , sector s at time t ,¹⁰ on $\log Spe_{pst}$ which captures the effect of specialisation in a given sector on productivity. The panel structure of the data (81 provinces, 659 industries and 6 years), allows me to introduce sector fixed-effects (γ_s) which capture sector-specific differences in the productivity that is irrespective of time (e.g., average labor productivity is higher in manufacturing sectors than in agriculture). Province-year fixed effects (γ_{pt}) can be considered as local wage indices after controlling for industry fixed-effects and localisation economies. Finally ε_{pst} is the error term representing unexplained productivity net of controls and fixed-effects.

The province-year fixed-effects estimated in the first step are then used as the dependent variable in the second step (Equation 2) and regressed on local characteristics that impact the productivity levels. To account for the local structure I use density ($\log Den_{pt}$), where Den_{pt} is the total number of employees (or population¹¹) in province p at time t (emp_{pt}) divided by land area ($Area_p$).¹²

The estimation of density could suffer from omitted variable bias. To address this concern, X_{pt} includes time-varying controls from economic geography literature (e.g. market potentials, diversity,

¹⁰I calculate the daily wage by dividing the average monthly salary by the number of days worked.

¹¹Depending on the mechanism studied, agglomeration effects can be measured using employment, population, or production. Since these three variables are highly correlated separate identification of their impact is not possible. I use employment (instead of the population) as it reflects better the magnitude of the local economic activity (Combes and Gobillon (2015)). Moreover, given that some of the local variables (such as diversity or specialisation) can only be constructed using employment data, it is more consistent to measure the employment density.

¹²Since Ciccone and Hall (1996), it is customary to measure the size of the local economy using density, which is the number of workers (or individuals) per unit of surface area. Although the number of employees can also be used directly, dividing over the land area is preferable as it addresses concerns due to heterogeneity in the spatial extent of the geographic units that are used. Moreover, the use of density addresses the concerns about the shape of the unit of analysis (due to the arbitrary administrative borders) which is known as the *modifiable areal unit problem* in the literature (Briant et al. (2010)). The β captures the total impact of local characteristics related to agglomeration economies rather than the magnitude of specific channels through which agglomeration forces operate. Moreover, it is the total of the net effects of density, which could be both positive due to agglomeration economies but also negative due to congestion costs.

human capital, road network, and more.) while Z_p , includes the time-invariant controls (e.g., land area, temperatures, geographic controls, etc.). Beyond their importance for identification of an unbiased estimate, their inclusion also makes it possible to quantify relative importance of each determinant in the overall productivity. Appendix Section B details the importance of these variables, their construction and data sources for each variable. γ_t is the year-fixed effect, which absorbs any temporal variations that affect the productivity of all provinces and sectors equally (e.g. productivity gains from technological progress). The use of time fixed-effects also addresses concerns due to the use of wages in current Turkish Lira and is more precise than using an arbitrary price deflator (Combes and Gobillon, 2015).¹³

This method is preferable for two reasons. First, doing a two-step estimation allows for estimating two separate error terms one for province-sector-year (ε_{pst}) and one for province-year (ε_{st}). This makes it possible, in a second-step, to tackle the endogeneity of density and other location characteristics without addressing the sector-specific endogeneity issues, such as specialisation ($\log Spe_{pst}$). Second, this procedure makes it possible to separately identify the localization economies (first-step) from those that are due to urbanization (second-step) as well. This is particularly important for policy formulation as it helps determine whether policy focus should be on further developing existing sectors or encouraging the arrival of new activities to the region.¹⁴

4.2 Estimation Issues

There are two sources of bias in the estimation of the equation above. First, workers may prefer to locate in large cities where wages are higher, creating a reverse causality problem between density and wages. Secondly, workers with higher ability may sort to larger cities which can bias positively the estimated elasticities. This section details these estimation issues and explains how I deal with each one.

4.2.1 Estimation Issue 1: Circular Causality

First identification issue is due to possible simultaneity between local characteristics and local wages. For example, employment areas receiving a positive technology shock may attract migrants, which would, in turn, increase the density of employment and further increase the wages. Such factors can generate

¹³Nominal wages are a better of measure for capturing differences in the productivity compared to real wages, which would be capturing differences in “standard of living” (Duranton, 2016). The estimation of local real wages requires considering the cost of living, specifically land prices. Given that such data is rarely available, nominal wages are used to have a consistent measure of productivity.

¹⁴The estimation can be done in a single-step. However, this estimation would be problematic because it does not allow computing the variance of local shocks. This makes it impossible to distinguish local shocks from purely idiosyncratic shocks at the industry-location level, which is vital with missing endowment variables. Furthermore, in a single-step estimation, the variance of local shocks has to be ignored when computing the covariance matrix of estimators. This can create significant biases in the standard errors for the estimated coefficients of aggregate explanatory variables (Moulton, 1990). To address this problem Combes et al. (2008) propose a two-step estimation strategy which both solves this issue and has the advantage of corresponding to a more general framework. As a robustness check, I run a single-step estimation and find very similar results. Results can be provided if requested.

a positive correlation between employment density and wages, causing an upward bias in the estimates. Other regressors such as market potential or human capital are also likely to be endogenous since they also depend on workers' and firms' location decisions.

Several instrumentation strategies have been proposed in the literature to address this endogeneity issue (Combes and Gobillon, 2015). Using historical instruments, as proposed by Ciccone and Hall (1996), is the most popular method and it builds on the hypothesis that historical values of population (or density) are relevant for today's levels as they are persistent over very long periods. The local outcomes of today (such as productivity or types of economic activities), however, are unlikely to be related to the economic outcomes a long time ago that probably affected the historical population. Following this strategy, I construct various instruments using Ottoman Empire population statistics of 1914 and the Turkish Republic's population census of 1927 and 1935.

Using these historical population numbers, I build variables that capture population densities and growth. The instruments are valid in the case of Turkey for a few reasons. First, it is unlikely that density levels from almost 100 years ago to be correlated with labor productivity today, as the Turkish economy went through a wide range of productivity shocks during this period. Successive wars between 1914 and 1923 had a significant impact on physical and human capital stock while disrupting industrial and agricultural production in most parts of the country (Pamuk, 2014). In addition, considerable population shifts took place between 1914 and 1924, which caused a dramatic reduction in the share of employment in the non-agricultural sector. The urban population was disproportionately affected by the decade-long wars and in their aftermath (Altug et al., 2008).

Economic government and policies have seen important changes as well. Following the transition from a multi-ethnic, multi-religious empire to a nation-state under a democratic and representation rule, the newly founded capital Ankara created policies that presented a contrast to the past. As discussed in Appendix Section A, Turkey went through important sectoral re-allocation and experienced a significant structural transformation. Massive public investment in human capital increased literacy rates from around 10% in 1923 to 90% in 2007 (Altug et al., 2008), while urban population increased dramatically following the large-scale mechanization in 1950s which freed up labor in rural areas.

I construct past domestic market potential using 1935 and 1945 population numbers, and past foreign market potential using the GDP levels of the main trading partners of Turkey in 1945 using the Maddison Project Database. In the main specification, I jointly use several instruments (instead of using only the 1914 urban population) for two reasons. First, having multiple instruments allows me to the instrument not only for employment density but also for the market potential, diversity, and even land area. Second, it makes it possible to carry out over-identification tests. Although I use multiple instruments in the main estimations, I also provide robustness tests using various single instruments.

I instrument density, market potential, and human capital. Although endogeneity of local character-

istics can be argued for almost all of them, I choose to instrument at most three of them simultaneously, as more than that would be extremely demanding in terms of identification power. I estimate these instrumented regressions either controlling for all or none of the non-instrumented variables and show that the results are consistent in both cases.¹⁵

4.2.2 Estimation Issue 2: Sorting by ability

The second identification problem is the possible correlation between density and worker characteristics. If workers' spatial distribution depends on their abilities, then the local productivity would also be affected by the differences in the composition of workers. In the case of sorting based on ability, the estimated impact of local variables would be inflated as they would be capturing productivity gains that are also due to differences in the composition of the workers' abilities that are observed and unobserved.

The sorting of more-skilled workers into larger cities is observed in the US and Europe (Combes et al. 2008; Baum-Snow and Pavan 2012; De la Roca and Puga 2017). In the case of the US, Baum-Snow and Pavan (2012) find sorting based on observables characteristics, yet none due to unobservable ones. In the case of France, however, Combes et al. (2008) show that controlling for observable skills is not enough to remove the bias. Contrasting the evidence from these developed countries, Combes et al., 2015 and Combes et al. (2020) find very weak sorting effects based on observables in China. This weak relationship makes them conclude that in the absence of sorting based on observables, sorting due to unobservable characteristics is unlikely. These results show that both the degree of bias and its direct channel (i.e., observable or/and unobservable) are specific to each context.

The most commonly used strategy is to use a panel of workers and estimate the parameter by comparing the same workers across several locations as suggested by Combes et al. (2008). However, such data is hard to find in developing countries. An alternative solution is to control for an extensive set of individual characteristics to take care of differences in the observable skills (Duranton 2016; Chauvin et al. 2017; Combes et al. 2020). If workers, however, sort across locations based on their unobservable abilities, this method is not enough to fully address the potential bias.

If such bias exists in Turkey, then a positive correlation between average wages and city characteristics could reflect a composition effect due to the over-representation of more able workers in some provinces. The inclusion of human capital as a control addresses concerns due to sorting based on observable characteristics. The sorting based on unobservables, however, would not be possible to address when using data aggregated at industry-province level. It is thus important to see whether such sorting exists in Turkey and if so, measure its size. To measure the potential bias due to sorting, I use the Household Labor Survey and exploit its individual-level dimension to account for sorting based on observable characteristics.

¹⁵For a similar argument see Combes et al. (2015, 2020).

5 Results

In this section, I start by estimating the elasticity of wages to density by addressing gradually each identification concern discussed previously (Section 5.1). Then I present results of a multivariate framework accounting for local characteristics that determine the local productivity (Section 5.2). I end the section with estimations that include infrastructure and other amenities (Section 5.3).

5.1 Density

I apply the two-step regression methodology where in the first-step I estimate province fixed-effects which are used as the dependent variable in the second step.¹⁶ Table 1 presents OLS results for the second step of the estimation. Each column corresponds to a different model using observations over 6 years for 81 provinces. I use weights that are proportional to the number of observations used to compute the LHS variable as it corrects for heteroskedastic error terms and thereby achieve a more precise estimation of the coefficients (Solon et al. (2015))¹⁷. Standard errors are clustered at province-level.

[Table 1 about here.]

Column 1 of Table 1 shows that density has an elasticity of 0.06. This means that in the cross-section of Turkish provinces between 2008 and 2013, doubling the worker density increases the average productivity by 4%.¹⁸ Concretely, if density of Iğdır (2.96, P25) were to increase to density of Mersin (11.3, P75), its productivity would increase by 14.7%.¹⁹

In the following columns, I address the omitted variable bias by adding a number of controls that are standard in the literature.²⁰ In Column 2, I add the land (surface) area of the province. The impact of land area is significant, which is in line with the literature. The positive coefficient suggests that, holding density constant, a 1% increase in the land area, increases the average productivity by 0.04%. The elasticity of density ($\beta = 0.06$) increases slightly compared to Column 1, although given the standard errors, I cannot reject the null hypothesis that they are equal. In Column 3, I add diversity control, which is also very significant. The elasticity of density drops to 0.051, suggesting that provinces with higher employment density also have a more diverse industry structure. In Column 4 and 5, I separately add controls to capture the domestic market potential (DMP) and the foreign market potential (FMP). Both controls are highly significant, and they reduce the elasticity of density.

¹⁶ Although the first-stage is not reported, all of the estimated province fixed-effects and specialisation are highly significant ($p < 0.001$). The estimated elasticity of log of specialisation is 0.065. This means that 1% increase in the specialisation of the local industry increases the average wages by 0.065%.

¹⁷ All of the results reported in the paper carry through without weights.

¹⁸ $2^{0.06} - 1 \approx 4\%$

¹⁹ This elasticity is the result of pooling observations across 6 years. However, as can be seen in Appendix Table D1, the coefficient remains stable when the regression is repeated separately by year. The yearly results also address any concerns related to pooling of 6 years of observations together and instrumenting with time-invariant historical variables. As can be seen from the results, coefficients remain highly stable and statistically significant, while the standard errors are small.

²⁰ It is important to note by including these controls, I make the implicit assumption that the controls are exogenous. Furthermore, these regressions ignore the endogeneity concerns due to reverse causality and individual unobserved heterogeneity. I address both concerns in the following sections separately.

In Column 6, I account for human capital by adding the share of the population with a university degree or higher. The control is highly significant, and its inclusion reduces the coefficient of density which is expected given the high correlation between density and human capital levels. This implies that about 20% of the relationship between wages and employment density is explained by worker characteristics and (possible) human capital spillovers. It is important to note, however, that human capital control is a share and is a semi-elasticity.

Finally, in Column 7, I include all of the controls to provide estimates that are comparable to the literature. When all controls are included, the coefficient of density is 0.0614 and is highly significant. This means that doubling the density of a location increases the average productivity by 4%, when all else is equal. While the land area remains significant, diversity turns insignificant which is in line with the literature (Combes and Gobillon, 2015). When regressed along with the domestic market potential, the foreign market potential becomes insignificant. This common problem is due to the high correlation between the two variables (≈ 0.93) and the lack of spatial variability of the foreign market demand given the way the variable is constructed.²¹ That is why, in the following sections, I work only with the domestic market potential. Lastly, the contribution of human capital remains powerful and significant.

5.1.1 Reverse Causality

The second important estimation issue is the reverse causality between density and wages. I address this concern and the potential bias it generates by implementing an IV strategy. This exercise is essential for the estimation of an unbiased elasticity because a good instrument would take care of both the reverse causality and the omitted variable bias addressing both concerns simultaneously.

My main identification strategy consists in implementing the 2-stage-least-squares (2SLS) estimation outlined in Section 4. I use the estimated province-year fixed-effects as dependent variable and include only density as an explanatory variable. In order to address the endogeneity of the main variables of interest, I use various historical instruments that are based on historical census data from 1914, 1927, 1935 and 1945. Separately, I instrument $\log Den_{pt}$ with population density in 1914 ($\log Den_{p1914}$), population density in 1927 ($\log Den_{p1927}$), employment density in 1935 ($\log Den_{p1935}$), employment density in 1945 ($\log Den_{p1945}$), and growth in population density between 1914 and 1927 ($DenGrowth_p$).

Figure 2 provides a visual representation of the first stage for density. It plots the density in 1914 in the horizontal axis against the average density in 2008 and 2013. Each observation in the figure corresponds to a province.

[Figure 2 about here.]

The figure shows that density in 1914 is a good predictor of the density today. I formally test the validity of each instrument by reporting the first-stage regression using the following equation:

²¹For more on the issue, see Redding and Venables (2004) and Combes et al. (2011).

$$\log Den_{pt} = \nu + \theta_1 \log Den_{p1914} + \gamma_t + \varepsilon_{pt}$$

Table 2 reports the coefficients from the first-stage regression. The coefficient θ , reported in the table represents the effect of the past densities on the current density levels. In line with the main specification, all of the regressions include time fixed-effects and errors are clustered at the province level.

[Table 2 about here.]

Each column presents the coefficients from regressions where I use a different instrument. In the first column of Table 2, I use the population density in 1914. The estimated coefficient is highly significant and around 1.17 which is similar to estimates reported in Combes et al. (2011). Specifically, an increase in the imputed past density by one percentage point leads to a 1.17 percentage point increase in the worker density between 2008 and 2013.

In the following columns, I repeat the exercise using densities from more recent years. All of the instruments regardless of the base year used or whether they capture employment or population density, pass the weak instrument test.²² As expected the F-statistic gets larger when I use densities from more recent periods. Results show that population densities at the beginning of the 20th century are good predictor for the employment density in 2008 and 2013 period due to the strong inertia of the urban hierarchy in Turkey. However, given the significant changes that took place between two periods, it is unlikely that they are correlated to labor productivity.

Table 3 presents the results for the estimation of Equation 2.²³ As explained previously, the parameter β_1 corresponds to the effect of employment density on the productivity levels of the provinces.

[Table 3 about here.]

In the first column, I report the elasticity obtained through OLS estimation. In the following columns, I present 2SLS results where the density is instrumented with lagged densities that are reported in the header of each column. In the final column, I use the change in the population density between 1914 and 1927 as an additional instrument.

Few results stand out. First, regardless of the instrument, the elasticities remain stable and highly significant around 0.056-0.06. This suggests that the OLS estimates suffer from a positive bias around 10%, which is in line with the literature (Combes and Gobillon, 2015). Second, all of the instruments have strong first-stages, proving to be good predictors. It also shows that the results are not dependent on the use of a specific instrument and are thus robust. Third, the standard errors are very small, indicating high precision of the estimates. Finally, the elasticities are very similar to that found in the

²²The number of observations in Columns 4 and 5 is lower due to differences in the number of provinces. Between 1923 and 2008, the number of provinces went from 57 to 81. The historical data in 1914 and 1927 were at district level, which allowed me to combine them according to the province boundaries in 2008. The data for 1935 and 1945, however, were only available at province-level. This made it impossible to distribute them according to the current number of provinces.

²³Appendix Table C1 reports the summary statistics for the second-stage estimation.

final column of Table 1 indicating that valid instruments can take care of both the bias due to reverse causality and missing variables.

Table 3 shows that once instrumented elasticity of productivity to density is between 0.056-0.06. This means that doubling the worker density increases the average productivity by 3.8 - 4%. This elasticity is comparable to those found in other countries that use similar specifications with aggregate data. It is similar to 0.06 found in Combes et al. (2008) for French employment areas over the period 1976-1998, 0.05 in Ciccone (2002) for the five largest EU-15 countries at the end of the 1980s, 0.06 found in Ciccone and Hall (1996) for American counties in 1988. These numbers are much higher than the estimates using individual level data and individual fixed effects, which find elasticities between 0.01 and 0.03 (Combes et al., 2008; Ahrend et al., 2017; De la Roca and Puga, 2017).

Compared to estimates in other developing countries, the elasticity of density is higher than 0.035 found in Ecuador (Matano et al., 2020) or 0.05 found in Colombia (Duranton, 2016), but lower than 0.09–0.12 found for India (Chauvin et al., 2017) or 0.10-0.12 found for China (Combes et al., 2015). Given the level of urbanization in Turkey, this elasticity fits precisely where it would be expected.²⁴

5.1.2 Sorting By Ability

If workers sort across locations based on their observed and unobserved abilities, the estimates can be biased up to 100% as shown in Combes et al. (2008) in the context of France. However, it is also possible for the bias not to exist if there is no correlation between individual characteristics (observed and unobserved) and local characteristics as found in China (Combes et al., 2017). This section explores whether sorting exists in Turkey and, if so, measure its importance.

Given that SGK data lacks the individual complement, it is impossible to test the existence of such bias and net out its effect if it exists. As a way to detect the presence of such bias, I use Household Labor Force Survey (LFS). The survey is conducted every quarter to measure the state of the economy. It covers around 500 000 individual observations annually, including all ages and both genders. It is sampled to be reflective of the Turkish population but also of the state of the economy, as it is used for calculating unemployment rate and other labor market measures.

The main shortcoming of this data, given the objective of this paper, is that it is aggregated at NUTS-2 region level. This means that the locations of individuals can be identified only at one of the 26 regions which is a shortcoming as smaller geographical units are better for capturing benefits of interactions that decay with distance (Combes and Gobillon, 2015).²⁵ Still, using larger local units may

²⁴Due to absence of panel data in the developing countries, these papers use individual-level cross-sectional data, which allows them to net out the endogeneity bias due to only the sorting effects based on observable characteristics. Thus their elasticities present gains that are due to agglomeration economies generated by the higher densities (and possible sorting effects based on unobservables).

²⁵26 regions correspond to the NUTS-2 level. These regions have different geographical and population sizes. NUTS-2 regions such as Istanbul (TR10), Izmir (TR31) and Ankara (TR51) are identical to the NUTS-3 provincial borders. Other regions are formed by combining multiple provinces. See Appendix for a map.

not be an important issue. According to Briant et al. (2010) using consistent empirical strategies (i.e. accounting for individual selection) largely reduces issues related to shape and size of the unit of analysis, and allow the estimation of unbiased estimates.

Different waves of the LFS are repeated cross-sections with no individual identifiers which makes it impossible to use individual fixed-effects. To examine the existence of a possible selection, I carry out two tests following follow Combes et al., 2015 and Combes et al. (2020).

As a first test, I examine the sources of spatial variations in wages. I begin by estimating the first step estimation including different sets of explanatory variables (location effects, individual characteristics, and firm characteristics) to understand their relative contribution to the log of the monthly wage. Table 4 reports the adjusted R^2 of each regression using observations from 2009.²⁶

Individual characteristics (i.e., education, age, and sex) alone explain 41% of the variations in individual wages. The explanatory power of firm characteristics (i.e., firm size) is 24%. Region effects (i.e., region dummies and specialisation) explain only 8%. These results suggest that individual characteristics are the main factors explaining individual wage disparities, followed by firm effects. The location effects and specialisation matter very little.

[Table 4 about here.]

These results reveal that three sets of effects are fairly orthogonal. Region effects and individual characteristics together explain 44% of the wage disparities while the sum of their individual R^2 is 0.49. Similarly, while region and firm effects explain 29% of the variation, the sum of their individual R^2 is 0.32. Finally, individual and firm characteristics explain only 45% of the differences in wages, the sum of their individual R^2 is equal to 0.65. These results suggest that differences in observed wages cannot be attributed to differences in the composition of the labor force or the type of firms present. The absence of correlation between the effect of individual characteristics and region dummies suggest that workers do not sort across regions according to their observable characteristics.

To further examine the absence of sorting in Turkey, I carry out a second exercise. If individuals sort across regions according to their abilities, some local variables, especially the density, should be correlated with individual observables such as education or occupation. It is possible to test this by estimating the region-year fixed-effects with and without individual characteristics in the first step of the estimation.²⁷ If workers sort across locations by their observable characteristics, these estimated fixed-effects should absorb them, and thus provide different results in the second step.

I combine multiple waves of the Household Labor Force survey for the period 2008 and 2013, and create a sample that matches the SGK data.²⁸ This leaves me with 412 137 individual observations over

²⁶Full estimation results are available if requested.

²⁷See Appendix Table E1 for the results.

²⁸Similar to the SGK sample, I keep all male and female workers, who are between 18-65 years of age, with positive income, employed in the private sector and affiliated to the social security system. There are a few reasons I apply this

the 6 year period which I use in a two-step procedure with individual level data similar to Combes et al. (2008). The procedure consists in estimating the following specification:

$$\log w_{irst} = \alpha + \delta \log Spe_{rst} + \phi Z_{it} + \gamma_s + \gamma_{rt} + \varepsilon_{irst} \quad (3)$$

$$\gamma_{rt} = \nu + \beta_1 \log Den_{rt} + \gamma_t + \varepsilon_{rt} \quad (4)$$

The first step estimation of equation 3 evaluates the impact of individual i 's wage at year t , w_{irst} , of region-time fixed-effects, γ_{rt} , for region r where worker i is employed at year t and region r 's specialisation in sector s where i is employed (for 88 Nace 2 industries), and a set of individual characteristics Z_{it} , such as age, age squared, sex, education (7 groups), occupation (39 ISCO 88 categories). In the second step, I use the estimated region-year fixed effect, γ_{rt} , as the dependent variable and regress it on region's employment density ($\log Den_{rt}$) and time fixed-effects, γ_t .²⁹

[Table 5 about here.]

Table 5 reports OLS and 2SLS results for the second step.³⁰ In Columns 1 and 2, I regress the region-year fixed-effects which are estimated in the first step which includes only region-year dummies and specialisation control. In Columns 3 and 4, I regress region-year fixed-effects, which are estimated in the first step where I include also a set of individual characteristics. While the first-step is weighted with survey weights, the second-step is weighted with the number of workers used to estimate the region year fixed-effects in the first-step. All regressions include year fixed-effects and errors are clustered at the region-level. In Columns 2 and 4, I instrument the current employment densities with the population density in 1914.

The elasticity of density remains stable across specifications, regardless of whether individual controls are included or not in the first step. Although inclusion of individual controls improves the precision of

condition. First, I drop workers in the informal sector to match it with SGK data in terms of coverage. Second most of the evidence in the literature use data in the formal sector (e.g., Combes et al., 2008; D'Costa and Overman, 2014; De la Roca and Puga, 2017). By focusing on formal employment allows me to provide numbers that are comparable with the literature. Still, in Appendix Section G, I provide estimates including also workers in the informal sector. I drop self-employed as it could mean a large set of occupations (e.g., street vendors, shop owners) in a context like Turkey and are also not included in the SGK data. Although not reported, I also tested the robustness of the results to make sure that they are not dependant on the sample selection. Inclusion of public sector employees slightly reduces the elasticity of density to 0.047-0.053, while the inclusion of self-employed does not change elasticities. Results are available if requested.

²⁹I measure density using the survey data for consistency. In Appendix Table E1, I test the robustness of my estimate by measuring employment density using the SGK employment data. Results are almost identical.

³⁰First-stage results are presented in the Appendix Section E.1. All the variables have the expected signs and are statistically significant. One result that is worth pointing is that the estimated elasticity of (log) specialisation is much smaller (0.0279) than the elasticities found in the main results using data aggregated at industry-location-year (0.06). The individual-level LFS data allow accounting for the individual characteristics (e.g. education, occupation) and estimating the effect of specialisation net of education and occupation. As sectoral choice and individual ability are highly correlated with industry characteristic, the estimated elasticity of specialisation using the aggregate data attributes part of the positive effect of ability on specialisation (and other variables that are measuring local characteristics). It is important to note, however, that part of the drop can also potentially be explained by the larger geographical scale. If localization benefits suffer from geographical decay, then it is reasonable to expect externalities due to specialisation to be weaker and the estimated coefficient to be smaller.

the estimates by reducing the standard errors, in both instances, the elasticities remain the same which shows that workers in Turkey do not sort across locations based on their observables.

Both tests indicate the absence of sorting based on observable characteristics across Turkish provinces. While the sorting can also be based on the unobservable characteristics of the individual, it is very unlikely for a sorting based on unobservables to take place in the absence of sorting based on observables (Combes et al., 2015, 2020). These results are in deep contrast with what is observed in developed countries where a significant fraction of the explanatory power of region effects arises from the sorting of workers (Combes and Gobillon, 2015). However, they are very similar to the findings of Combes et al., 2015 and Combes et al. (2020) for China.

The similarity between these findings and those found in China (Combes et al., 2015, 2020) could be indicative of some significant differences between developed and developing countries in terms of sources of productivity differences. While sorting based on individual abilities seems to be an important determinant of spatial wage differences in developed countries, this pattern does not seem to hold in Turkey or China. To explain the lack of sorting in China, Combes et al., 2015 argue that mobility restrictions due to the Hukou system could be preventing workers to sort across urban areas based on their ability. Such mobility restrictions do not exist in Turkey. On the other hand, Turkey has seen experienced massive rural-urban migration since the 1950s. These migrations waves were triggered by the mechanization of agriculture across Anatolia and ethnic conflict that has hit the southeast of Turkey since 1985. These massive migration moves were directed to bigger cities but mainly to the three big cities. For instance, between 1950 to 2008, Istanbul’s population increased from 1.2 to 18 million. The arrival of such big waves of low-skilled workers originating from agricultural regions may have broken the link between urban externalities and the sorting.

5.2 Multivariate Approach: A Unified Framework

As discussed in Section 1, spatial wage disparities can be explained in three broad categories (i.e., skills, endowments and interactions). Combes et al. (2008) propose a “unified framework” which includes all of these explanations to have a sense of the magnitudes of each contributing factor. Understanding the contribution of each factor is especially important to inform policy.

I estimate the Equation 2, including a set of controls that capture all of the explanations. Naturally, this exercise is demanding in terms of data and requires instrumenting of multiple variables simultaneously. As the exogeneity of economic geography variables is debatable, one needs to be cautious when including them. In my analysis, I introduce five local variables (Density, Domestic Market Potential, Human Capital, Land Area and Diversity). Similar to Combes et al. (2020) I instrument at most three of them simultaneously (*Density, Domestic Market Potential, Human Capital*), as more than that is demanding in terms of identification power. I start with estimating only with these instrumented variables

and then include the other non-instrumented variables (*Land Area, Diversity*) and show that the results are consistent in both cases.

Specifically, I instrument $\log Den_{p,t}$ with population density in 1914 ($\log Den_{p1914}$) and growth in population density between 1914 and 1927 ($DenGrowth_p$); domestic market potential ($\log DMP_{pt}$) with domestic market potential in 1945 ($\log DMP_{p1945}$); and human capital ($HumanCapital_{pt}$) with number of enrolled male students in 1927 ($EnrolledMale_{p1927}$).³¹ In order to test the statistical relevance of these instruments, I report Cragg–Donald F-Statistic weak instrument test. I also report Shea’s partial R^2 to examine how much of the variation of the instrumented variables are explained by instruments, once potential inter-correlations among instruments have been accounted for. Finally, the Hansen J-Statistic tests over-identifying restrictions.

[Table 6 about here.]

Table 6 presents the 2SLS results.³² For comparability, I start by presenting the OLS and 2SLS results where density is the only explanatory variable. Column 1 presents OLS results for all provinces. Column 2 presents OLS results where 3 provinces for which instrument is not available are excluded. The elasticity hardly changes. Column 3 presents 2SLS results where density is instrumented with population density in 1914 and growth in population density between 1914 and 1927. In Column 4 and 5, I add domestic market potential, which I instrument using the domestic market potential in 1945. Introduction of this additional variable does not change the elasticity of density. In Columns 6 and 7, I include the human capital control which lowers the magnitude of both density and domestic market potential. This points to the relatively unequal distribution of the share of high skilled individuals, and a relatively strong correlation between employment density and human capital (Pearson’s $R \approx 0.44$). In addition to the instruments used in the previous regressions, I add the number of enrolled male students in 1927 as an additional instrument. This positive and significant effect is expected as it captures both the private gains due to skills and the externalities generated by the presence of higher skilled individuals in an agglomeration.³³

In Columns 8 and 9, I further control for the land area and diversity of the local economic activity. Although the former can be considered exogenous to density, the latter is correlated. The coefficient of the land area is statistically significant, and suggests that for a given level of density, an increase in the land area increases the average productivity. If, for example, the land size of a province doubles, the

³¹I have a broader set of possible instruments, and I experimented with multiple combinations. I have a broader set of possible instruments, and I experimented with many combinations, all of which yielded broadly consistent results with one another and with the OLS. I report estimations using the same sets of instruments to allow for reliable comparisons. I also try to be restrictive about the number of instruments and use just enough to carry out over-identification tests. Results using other sets of instruments are available if requested.

³²Appendix H1 table reports the first-stage results.

³³The coefficient on the share of high skilled workers will also capture complementarities between skilled and unskilled labor in the production function. Also, as more educated workers flock to cities with higher wages, it also adds to the identification issues. Controlling for human capital make progress towards the identification of the true elasticity of density but also make it possible to compare it with the findings in the literature.

wages increase around 3%. The diversity, on the other hand, is insignificant which is quite common in the literature (Combes and Gobillon, 2015).

Finally, I account for endowments. As discussed earlier, many productive endowments (such as airports, high-speed train lines, highways) can increase wages. However, given the endogeneity concerns in using such controls, I consider only four (exogenous) endowment variables that are related to the geography and thus are less concerning in terms of endogeneity. In Columns 10 and 11, I include controls to account for differences in length of seashores within the provincial boundaries (*Shores*), access to sea coast (*Coast*), mean annual temperature (*Climate*) and presence of rivers (*Rivers*).³⁴ Compared to the previous columns, the inclusion of these controls does not impact the other coefficients. The coefficient of *Coast* is positive and significant, suggesting that having access to the shore improves productivity.³⁵ While the *Shores* and *Rivers* do not seem to have any effect, the *Climate* seems to decrease productivity. Overall, the inclusion of these variables do not increase the explanatory power of the regressions (R^2) which is already high.

The last column (11) is my preferred specification, as it is the most comprehensive one. It includes controls that account for skills-based endowments (Human Capital), between-industry interactions (*Density*, *DMP*, *Human Capital* and *Land Area*) and amenities (*Shore*, *Coast*, *Climate* and *Rivers*). Density, domestic market potential, and human capital are instrumented with long-lagged variables. The results pass all the relevant statistical test, and the model has high explanatory power.

The elasticity of density is 0.057, which is exactly the same as those found in Table 3. The domestic market potential remains positive and highly significant. The estimated coefficient (0.1) is a little less than the double of density, suggesting that having access to other markets is the most important determinant of the productivity differences across Turkish provinces. If the market potential of a province doubles (e.g., employment density doubles in all other provinces), the wages increase by 6.5%. This number is more than the triple of the 0.02 found for France in Combes et al. (2008), but smaller than 0.13-0.22 found for China in Combes et al. (2020).³⁶ Once instrumented, human capital spillovers becomes statistically insignificant. When the same regression is estimated without density, the coefficient on human capital is high and significant. Given the high correlation between the two variables (Pearson's $R \approx 0.44$), it is possible that the effect of human capital is captured by density.³⁷

³⁴I drop diversity as it is insignificant and generates unnecessary endogeneity.

³⁵It is important to note that this variable captures the walking distance to the closest seashore. It does not imply having access to a port, which would be highly endogenous.

³⁶The impact of market size on wages has been studied for China (Au and Henderson, 2006), India (Lall et al., 2004), and Colombia (Duranton, 2016). These studies also find that market size has a larger effect than in developed countries.

³⁷The mixed results regarding human capital externalities is not unusual for developing countries (Duranton, 2016). Still, to check whether these results are driven by any of the choices made when estimating the regression, I make several tests: First, I instrument human capital with alternative instruments such as the number of enrolled female students in 1927, schools to population ratio in 1927, schools to students ratio in 1927, share of high school in 1985. Second, in case the share of educated population in province population is not used with an appropriate functional form, I use the log of human capital variable. Third, I use alternative human capital measures which go from the more restrictive (university degree) to the least restrictive (post-secondary education). In all instances, results and elasticities for all variables hold while the human capital remain positive yet statistically insignificant. Results are available if requested.

5.3 Additional results: Infrastructure and other amenities

In this section, I extend the number of controls used in the previous section and add three sets of variables that can impact the wages: transport infrastructure, cultural and additional set of climatic amenities. The infrastructure amenities (e.g., road infrastructure, train-lines or airports) can improve the productivity by affecting the growth of urban areas, increasing trade and lowering cost of transportation (Redding and Turner, 2015). Although better infrastructures can improve overall productivity and increase wages, it can also reduce them through improving market access (thus lowering the prices of inputs and goods). The final effect on the wages is therefore, ambiguous. The literature on the effects of infrastructure is limited, mainly due endogeneity issue making it difficult to establish a causal effect.

Cultural amenities such as cinema halls, theaters or parks, can also impact the wages as they may increase the willingness of consumers to pay for land and thus imply higher local land rents (Roback, 1982). When the local prices increase, firms use relatively less land, which in turn can decrease the marginal product of labor, especially if the latter and land are not perfect substitutes. Similarly, amenities related to climate can also have a similar effect as it may increase the cost of land and living through higher housing prices and lower wages (Glaeser and Gottlieb, 2009).

[Table 7 about here.]

Table 7 reports the OLS results where I augment the model in Section 5.2 with additional controls. In all of the regressions I control for market potential, human capital, geographic endowments, land area and include time fixed-effects.

Column 1 replicates the results from Table 6 for comparison. In Column 2, I account for the length of provincial road network (*Roads*) and village roads (*Village Roads*). The results suggest that a denser road network does not impact the wages.³⁸ As discussed earlier, better road networks can impact both positively and negatively the average wages. The non-significance of the results could be due to two opposing effects canceling each other out, and does not prove the absence of an effect.³⁹ It should also be noted that these coefficients capture the effects that remain when controlling for domestic market potential.

In Column 3, I control for cultural amenities and health facilities. The number of cinema halls and libraries have the expected negative sign, although only the former is weakly significant (at 20% significance level). Hospitals, on the other hand, have a positive sign yet is insignificant. The interpretation of this sign is also should be done with caution. While better hospitals can increase the demand for the location, it can also be the consequence of higher levels of local income.⁴⁰

³⁸It should be noted that provincial and village roads are inferior to highways. While village roads are correlated with lower densities and rural economic structure, the relationship between provincial roads and density follows an inverse U-shape. That is why attempts to establish a linear relationships should be addressed with caution. Although not reported here, the highway network, which was very limited in the period of analysis, does not change results and remain insignificant.

³⁹For more on the issue, see Duranton (2016).

⁴⁰Although not reported, I also tested for other amenities such as the number of theaters, museums, doctors, clinics, the number of beds in hospitals, and more.

In Column 4, I add controls to capture climate-related amenities following Chauvin et al. (2017). I account for average temperatures in January and June, and their difference from ideal temperatures of 21.11 Celcius.⁴¹ I also control for the average number of sunny and rainy days in a month. The results suggest that Turks seem to get higher wages when they live in less temperate climates, although the magnitude is small. Despite significant differences in climate between Turkish provinces, it does not seem to matter for productivity or average wages.⁴² This result is not surprising given that Turkey's economic divide goes from west-to-east, and other controls in the regression capture it. Finally, the provinces with a higher number of average sunny days seem to enjoy higher productivity levels. This result is possibly driven by the fact that Turkish provinces located on the Mediterranean coast have higher average incomes.

Given the endogeneity concerns and the difficulty in instrumenting, these results should be viewed as robustness checks for the main findings. The absence of significant and robust effects are potentially due to the high correlation between density (and the other controls that are included) with the amenities. The spatial inequalities are present in every measurable metric and thus are highly correlated with wages and average income in Turkey. This collinearity may prevent the appropriate identification of the effects of these amenities. For similar results found in the context of Colombia, Duranton (2016) argues that the collinearity is unlikely to be problematic when the standard errors for the amenities are small, which means that they are fairly precisely estimated. If true, these results indicate the weak relationship between productivity (wages) and amenities in Turkey which corroborate findings in other developing countries (Duranton, 2016; Chauvin et al., 2017).

6 Conclusion

This paper contributes to the literature on the agglomeration economies by providing evidence from Turkish provinces. Turkey is an excellent example of a developing country that has experienced fast urbanization and has a high share of the population living in urban areas. Yet, the findings of this paper has implications that go beyond Turkey. In addition to providing the first estimates on the determinants of spatial differences in productivity in Turkey, my findings also contribute to the broader knowledge base about agglomeration economies in developing countries.

Using a novel administrative dataset, employing various panel data techniques and instruments based on historical data, I find a positive and causal effect of density on productivity in Turkey. The estimated elasticity of 0.057-0.06 is higher than those estimated across U.S. and Europe, around those found for Colombia and Brazil, and smaller than those found for China and India. Consistent with the previous

⁴¹The choice of 21.11 Celcius represents the middle ground between 18 and 24 degrees which is considered to be the ideal temperature for human comfort. Still, I check the robustness of this finding by using the similar measures obtained from Global Climate Data. Results can be provided if requested.

⁴²This finding is in line with the literature (for Colombia (Duranton, 2016), or China and India (Chauvin et al., 2017) which shows that amenities (climate-related but also others) do not matter for developing countries.

literature, I find a positive and significant effect for the market access, which is much stronger than those found in developed countries.

Remarkably, I do not find evidence for the sorting of workers across provinces based on their observed abilities. This result contrasts what has been found for developed countries while echoing the findings for China. Put together, these findings hint that the models and stylized facts documented for cities in the developed countries may not apply fully to developing countries, thus requiring the extension of current models to match the realities in developing country contexts. A better understanding of the factors that differentiate sources and gains associated with urbanisation in developed countries from developing ones are essential missing pieces in the literature, and thus remains high on my research agenda.

References

- Ahrend, Rudiger, Emily Farchy, Ioannis Kaplanis, and Alexander C. Lembcke. “What Makes Cities More Productive? Evidence From Five Oecd Countries on the Role of Urban Governance.” *Journal of Regional Science* 57, 3: (2017) 385–410.
- Altug, Sumru, Alpay Filiztekin, and Sevket Pamuk. “Sources of Long-Term Economic Growth for Turkey.” *European Review of Economic History* 12, 3: (2008) 393–430.
- Atalik, Gündüz. “Some effects of regional differentiation on integration in the European community.” *Papers of the Regional Science Association* 69, 1: (1990) 11–19.
- Atesagaoglu, Orhan Erem, Ceyhun Elgin, and Oguz Oztunali. “TFP growth in Turkey revisited: The effect of informal sector.” *Central Bank Review* 17, 1: (2017) 11–17.
- Atiyas, I, and O Bakis. “Aggregate and sectoral TFP growth in Turkey: a growth accounting exercise.” *Iktisat, Isletme ve Finans* 29, 341: (2014) 9–36.
- Au, Chun Chung, and J. Vernon Henderson. “Are Chinese cities too small?” *Review of Economic Studies* 73, 3: (2006) 549–576.
- Balkan, Binnur, and Semih Tumen. “Immigration and prices: Quasi-experimental evidence from Syrian refugees in Turkey.” *Journal of Population Economics* 29, 3: (2016) 657–686.
- Baslevant, Cem, and Ozlem Onaran. “Are married women in Turkey more likely to become added or discouraged workers?” *Labour* 17, 3: (2003) 439–458.
- . “The Effect of Export-Oriented Growth on Female Labor Market Outcomes in Turkey.” *World Development* 32, 8: (2004) 1375–1393.
- Baum-Snow, Nathaniel, and Ronni Pavan. “Understanding the city size wage gap.” *Review of Economic Studies* 79, 1: (2012) 88–127.
- Bolt, Jutta, and Jan Luiten van Zanden. “The Maddison Project: Collaborative research on historical national accounts.” *Economic History Review* 67, 3: (2014) 627–651.
- Boratav, Korkut. *Türkiye İktisat Tarihi 1908-2015*. Istanbul: Imge Kitabevi Yayinlari, 2017.
- Briant, Anthony, Pierre-Philippe Combes, and Mirene Lafourcade. “Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations?” *Journal of Urban Economics* 67, 3: (2010) 287–302.
- Celbis, Mehmet Güney, and Denis de Crombrughe. “Internet infrastructure and regional convergence: Evidence from Turkey.” *Papers in Regional Science* 97, 2: (2018) 387–409.

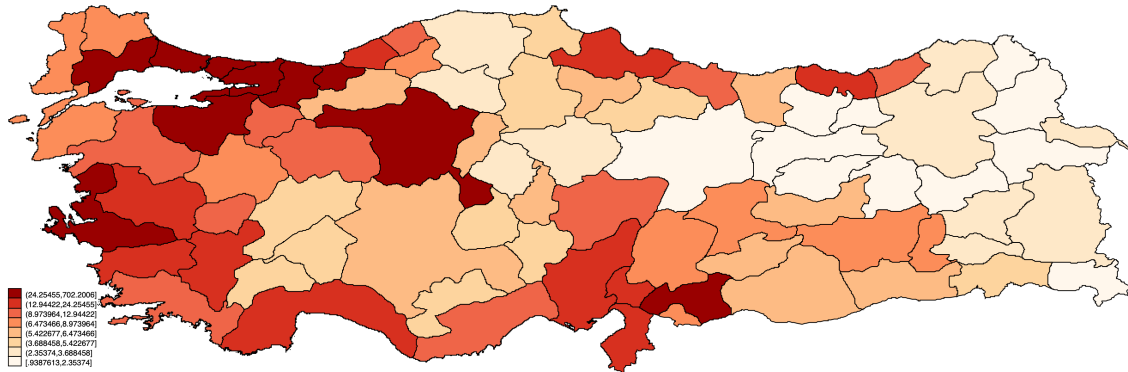
- Celebioglu, Fatih, and Sandy Dall'erba. "Spatial disparities across the regions of Turkey: An exploratory spatial data analysis." *Annals of Regional Science* 45, 2: (2010) 379–400.
- Chauvin, Juan Pablo, Edward Glaeser, Yueran Ma, and Kristina Tobio. "What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States." *Journal of Urban Economics* 98: (2017) 17–49.
- Ciccone, Antonio. "Agglomeration effects in South-South FDI." *European Economic Review* 46: (2002) 213–227.
- Ciccone, By Antonio, and Robert E Hall. "Productivity and the Density of Economic Activity." *The American Economic Review* 86, 1: (1996) 54–70.
- Combes, Pierre-Philippe. "Economic Structure and Local Growth: France, 1984-1993." *Journal of Urban Economics* 47, 3: (2000) 329–355.
- Combes, Pierre-Philippe, Sylvie Démurger, and Shi Li. "Migration externalities in Chinese cities." *European Economic Review* 76: (2015) 152–167.
- . "Productivity gains from agglomeration and migration in the people's Republic of China between 2002 and 2013." *Asian Development Review* 34, 2: (2017) 184–200.
- Combes, Pierre-Philippe, Sylvie Démurger, Shi Li, and Jianguo Wang. "Unequal migration and urbanisation gains in China." *Journal of Development Economics* 142, 102328.
- Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. "Spatial wage disparities: Sorting matters!" *Journal of Urban Economics* 63, 2: (2008) 723–742.
- Combes, Pierre-Philippe, and Laurent Gobillon. *The Empirics of Agglomeration Economies*, volume 5. Elsevier B.V., 2015, 1 edition.
- Combes, Pierre-Philippe, Miren Lafourcade, JF Jacques Francois Thisse, and Jean Claude Toutain. "The rise and fall of spatial inequalities in France: A long-run perspective." *Explorations in Economic History* 48, 2: (2011) 243–271.
- Cosar, Kerem, and Banu Demir. "Domestic road infrastructure and international trade: Evidence from Turkey." *Journal of Development Economics* 118: (2016) 232–244.
- Coulibaly, Souleymane, Uwe Deichmann, and Somik Lall. "Urbanization and Productivity: Evidence from Turkish Provinces over the Period 1980-2000." *World Bank* , August: (2007) 32.
- D'Costa, Sabine, and Henry G. Overman. "The urban wage growth premium: Sorting or learning?" *Regional Science and Urban Economics* 48: (2014) 168–179.

- De la Roca, Jorge, and Diego Puga. "Learning by working in big cities." *Review of Economic Studies* 84, 1: (2017) 106–142.
- Duranton, Gilles. "Agglomeration effects in Colombia." *Journal of Regional Science* 56, 2: (2016) 210–238.
- Duranton, Gilles, and Diego Puga. "Nursery cities: Urban diversity, process innovation, and the life cycle of products." *American Economic Review* 91, 5: (2001) 1454–1477.
- . "Micro-Foundations of urban agglomeration economies." In *Handbook of Regional and Urban Economics*, edited by J.Vernon Henderson, and JF Jacques Francois Thisse, Elsevier, 2004.
- Fally, Thibault, Rodrigo Paillacar, and Cristina Terra. "Economic geography and wages in Brazil: Evidence from micro-data." *Journal of Development Economics* 91, 1: (2010) 155–168.
- Gezici, Ferhan, and Geoffrey J.D. Hewings. "Regional convergence and the economic performance of peripheral areas in Turkey." *Review of Urban and Regional Development Studies* 16, 2: (2004) 113–132.
- Glaeser, Edward, and J. Vernon Henderson. "Urban economics for the developing World: An introduction." *Journal of Urban Economics* 98: (2017) 1–5.
- Glaeser, Edward L., and Joshua D. Gottlieb. "The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States." *Journal of Economic Literature* 47, 4: (2009) 983–1028.
- Harris, Chauncy D. "The market as a factor in the location of industry in the United States." *Annals of the Association of American Geographers* 44, 4: (1954) 315–348.
- Head, Keith, Thierry Mayer, and John Ries. "The erosion of colonial trade linkages after independence." *Journal of International Economics* 81, 1: (2010) 1–14.
- Hering, Laura, and Sandra Poncet. "Market access and individual wages: Evidence from China." *Review of Economics and Statistics* 92, 1: (2010) 145–159.
- Ismihan, Mustafa, and Kivilcim Metin-Ozcan. "Productivity and Growth in an Unstable Emerging Market Economy: The Case of Turkey, 1960-2004." *Emerging Markets Finance and Trade* 45, 5: (2009) 4–18.
- Jacobs, Jane. *The Economy of Cities*. New York: Vintage Books, 1969.
- Karpat, Kemal H. *Ottoman Population 1830-1914*. Madison, WI: University of Wisconsin Press, 1985.
- Kirdar, Murat G., and Sirin D. Saracoglu. "Migration and regional convergence: An empirical investigation for Turkey." *Papers in Regional Science* 87, 4: (2008) 545–566.

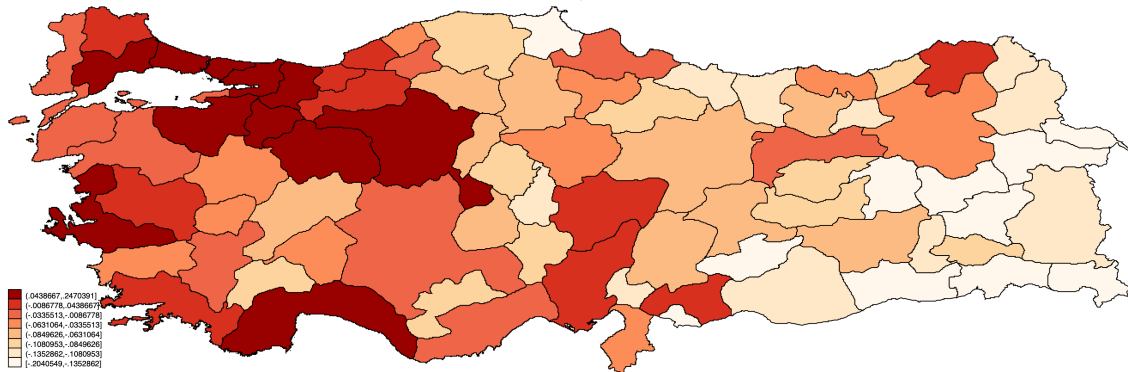
- Krueger, Anne O., and Baran Tuncer. "Growth of factor productivity in Turkish manufacturing industries." *Journal of Development Economics* 11, 3: (1982) 307–325.
- Krugman, P., and A. J. Venables. "Globalization and the Inequality of Nations." *The Quarterly Journal of Economics* 110, 4: (2006) 857–880.
- Krugman, Paul R. "Scale Economies, Product Differentiation, and the Pattern of Trade." *American Economic Review* 70, 5: (1980) 950–959.
- Lall, Somik V., Zmarak Shalizi, and Uwe Deichmann. "Agglomeration economies and productivity in Indian industry." *Journal of Development Economics* 73, 2: (2004) 643–673.
- Luca, Davide. "Votes and Regional Economic Growth: Evidence from Turkey." *World Development* 78: (2016) 477–495.
- Lucas, Robert E. "On the mechanics of economic development." *Journal of Monetary Economics* 22, 1: (1988) 3–42.
- Matano, Alessia, Moisés Obaco, and Vicente Royuela. "What drives the spatial wage premium for formal and informal workers ? The case of Ecuador." *Journal of Regional Science* 1–25.
- Metin, Karadag, A Onder Ozlem, and Deliktas Ertugrul. "Growth of factor productivity in the Turkish manufacturing industry at provincial level." *Regional Studies* 39, 2: (2005) 213–223.
- Moretti, Enrico. "Estimating the Social Return to Education: Evidence from Repeated Cross-Sectional and Longitudinal Data.", 1999.
- Moulton, Brent R. "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units." *The Review of Economics and Statistics* 72, 2: (1990) 334.
- OECD. *Regions and Cities at a Glance*. Paris, 2018.
- . "Global Cities." Technical report, Organisation for Economic Co-operation and Development, Paris, 2020.
- Onder, A, E Deliktas, and A Lenger. "Efficiency in the Manufacturing Industry of Selected Provinces in Turkey: A Stochastic Frontier Analysis." *Emerging Markets Finance and Trade* 39, 2: (2003) 98–113.
- Pamuk, Sevket. "Agricultural Output and Productivity Growth in Turkey since 1880." In *Agriculture and Economic Development in Europe since 1870*, edited by P. Lains, and V Pinilla, London, 2008. Routledge edition.
- . *Turkiye'nin 200 yÄ±ll ± kiktisaditarihi*. Istanbul : *TurkiyeIsBankasiKulturYayinlari*, 2014, 2.editionedition.

- Rauch, James E. "Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities." *Journal of Urban Economics* 34, 3: (1993) 380–400.
- Redding, S, and AJ Venables. "Economic Geography and International Inequality." *Journal of international Economics* 62, 1: (2004) 53–82.
- Redding, Stephen J., and Matthew A. Turner. "Transportation Costs and the Spatial Organization of Economic Activity." In *Handbook of Regional and Urban Economics, Elsevier, 2015, volume 5, 1339–1398.*
- Roback, Jennifer. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90, 6: (1982) 1257–1278.
- Sakalli, Seyhun Orcan. "Secularization and Religious Backlash : Evidence from Turkey.", 2019.
- Solon, Gary, Steven J Haider, and Jeffrey Wooldridge. "What Are We Weighting For." *The Journal of Human Resources* 50, 2: (2015) 301–316.
- Temel, Tugrul, Aysit Tansel, and Peter J. Albersen. "Convergence and Spatial Patterns in Labor Productivity: Nonparametric Estimation for Turkey.", 1999.
- Tumen, Semih. "The Economic Impact of Syrian Refugees on Host Countries: Quasi-Experimental Evidence from Turkey." *American Economic Review* 106, 5: (2016) 456–460.
- Turkstat. "Turkstat Regional Statistics (database).", 2019.
- World Bank. "Rise of the Anatolian Tigers." *Technical Report 87180, The World Bank Group, Washington DC, 2015.*

Figure 1: Spatial Differences
Employment Density (2013)



Productivity (2013)



GDP per capita in USD (2017)

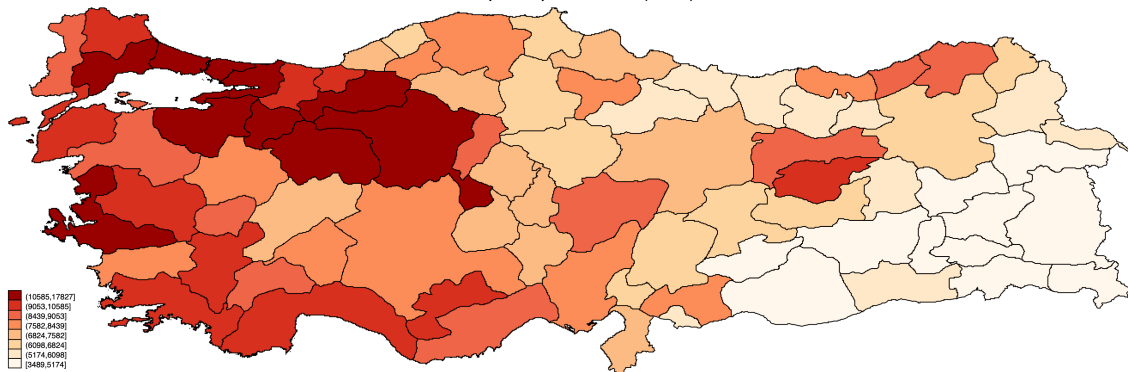


Figure 2: Instrumental Variable Relevance

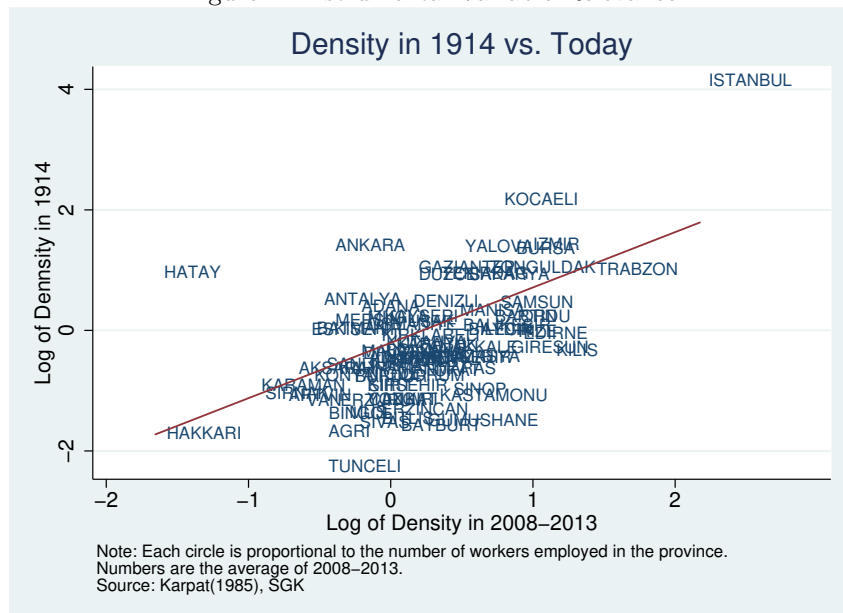


Table 1: OLS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Density	0.0624 (0.005)***	0.0700 (0.004)***	0.0506 (0.005)***	0.0608 (0.002)***	0.0449 (0.007)***	0.0504 (0.005)***	0.0614 (0.007)***
Land Area		0.0433 (0.018)**					0.0355 (0.010)***
Diversity			0.0767 (0.025)***				0.0170 (0.020)
DMP				0.0790 (0.014)***			0.0950 (0.023)***
FMP					0.2805 (0.086)***		-0.0592 (0.100)
Human Capital						5.7470 (1.338)***	3.0584 (0.869)***
N	486	486	486	486	486	486	486
Adj R2	0.83	0.86	0.85	0.88	0.85	0.87	0.94

The table reports OLS estimates for the impact of employment density on average productivity. The unit of observations are provinces. Regressions are weighted by total employment in year. Standard errors are clustered at province-level. Data source: Turkstat, SGK

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: First Stage Regressions, 2008-2013

	(1)	(2)	(3)	(4)	(5)
	LnDen	LnDen	LnDen	LnDen	LnDen
Density1914	1.4799 (0.197)***				
Density1927		1.1646 (0.110)***			
Density1935			1.7960 (0.171)***		
Density1945				0.0549 (0.003)***	
DenGrowth					0.0077 (0.003)***
N	468	486	342	378	468
Adj R2	0.76	0.80	0.81	0.85	0.46

The table reports OLS estimates for the impact of employment density in on average productivity. The unit of observations are provinces. The number of observations in Columns 3 and 4 is lower due to differences in the number of provinces. See footnote 21 for details. Regressions are weighted by total employment in year. Standard errors are clustered at province-level. Data source: SGK, Karpat(1985)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: 2SLS Results for 2008-2013

	OLS	Den 1914	Den 1927	Den 1935	Den 1945	Den Growth
Density	0.0624 (0.005)***	0.0560 (0.005)***	0.0582 (0.004)***	0.0547 (0.004)***	0.0537 (0.004)***	0.0632 (0.007)***
N	486	468	486	342	378	468
KP F-Stat		56.70	112.19	110.94	425.13	8.91

The table reports 2SLS estimates for the impact of employment density in on average productivity. The excluded instruments are reported at the header of each column. The unit of observations are provinces. The number of observations in Columns 3 and 4 is lower due to differences in the number of provinces. See footnote 35 for details. Regressions are weighted by total employment in year. Standard errors are clustered at province-level. Data source: SGK, Karpat(1985)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Explanatory Power of Various Sets of Variables

Individual characteristics	0.41
Firm characteristics	0.24
Region effects	0.08
Region effects and individual characteristics	0.44
Region effects and firm characteristics	0.29
Individual and firm characteristics	0.45
All three sets	0.48
N	59078

Notes: Table presents Adjusted R-squares for individual wage regressions using data for 2008. Region effects include both region dummies and the specialisation variable. Data source: Household Labor Force Survey (Turkstat), Karpat(1985)

Table 5: Second Stage Regressions, 2008-2013

	No Controls		Individual Controls	
	OLS	2SLS	OLS	2SLS
Density	0.0616 (0.003)***	0.0560 (0.007)***	0.0534 (0.003)***	0.0512 (0.005)***
N	156	156	156	156
KP F-Stat		66.85		66.85

The table reports OLS and 2SLS estimates for the impact of employment density in on average productivity. The unit of observations are NUTS-2 Regions. Regressions are weighted by total employment in year. Standard errors are clustered at region-level. Data source: Household Labor Force Survey (Turkstat), Karpas(1985)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: 2SLS Results for 2008-2013

	OLS	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Density	0.0624 (0.005)***	0.0623 (0.005)***	0.0576 (0.004)***	0.0609 (0.002)***	0.0571 (0.003)***	0.0483 (0.002)***	0.0450 (0.004)***	0.0594 (0.005)***	0.0672 (0.010)***	0.0560 (0.004)***	0.0574 (0.008)***
DMP		0.0792 (0.014)***	0.0821 (0.017)***	0.0816 (0.015)***	0.0821 (0.017)***	0.0816 (0.015)***	0.0791 (0.017)***	0.0878 (0.016)***	0.0983 (0.019)***	0.1102 (0.019)***	0.1165 (0.022)***
Human Capital				5.9888 (0.530)***			7.1728 (1.831)***	2.6207 (0.881)***	0.5091 (2.091)	2.5828 (0.885)***	1.8959 (2.237)
Land Area								0.0387 (0.010)***	0.0517 (0.018)***	0.0422 (0.007)***	0.0447 (0.011)***
Diversity								0.0175 (0.021)	0.0105 (0.023)		
Shores										0.0021 (0.002)	0.0029 (0.003)
Coast										0.0599 (0.018)***	0.0680 (0.021)***
Climate										-0.0039 (0.001)***	-0.0061 (0.002)***
Rivers										0.0000 (0.000)	0.0000 (0.000)
N	486	468	468	468	468	468	468	468	468	468	468
r ²	0.83	0.83	0.82	0.88	0.88	0.92	0.92	0.94	0.94	0.95	0.95
Cragg-Donalds Stat			1161.06		443.73		131.58		68.02		33.83
P-value Hansen Test			0.277		0.195		0.145		0.336		0.495
Shea's Partial(Density)			0.725		0.723		0.575		0.566		0.502
Shea's Partial(DMP)					0.791		0.702		0.512		0.582
Shea's Partial(HC)							0.691		0.619		0.673

The table reports OLS and 2SLS estimates for the impact of employment density and other controls on average productivity. The unit of observations are provinces. Regressions are weighted by total employment in year. Standard errors are clustered at province-level. Data source: Turkstat, SGK, Karpat(1985)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: OLS Results for 2008-2013: Infrastructure and Amenities

	(1)	(2)	(3)	(4)	(5)
Density	0.0573 (0.004)***	0.0596 (0.003)***	0.0581 (0.006)***	0.0608 (0.003)***	0.0569 (0.006)***
Roads(KM)		0.0206 (0.028)			0.0116 (0.021)
Village Roads(KM)		-0.0142 (0.017)			0.0052 (0.019)
Cinema Halls			0.0085 (0.007)		0.0055 (0.007)
Hospitals			-0.0088 (0.016)		0.0107 (0.016)
Libraries			-0.0011 (0.018)		-0.0166 (0.013)
January				0.0011 (0.002)	0.0015 (0.002)
June				0.0048 (0.002)**	0.0055 (0.003)**
Sunny Days				0.0169 (0.005)***	0.0196 (0.007)**
Average Rain				-0.0001 (0.000)	-0.0000 (0.000)
Market Potential	Yes	Yes	Yes	Yes	Yes
Human Capital	Yes	Yes	Yes	Yes	Yes
Geographic Endowments	Yes	Yes	Yes	Yes	Yes
Land Area	Yes	Yes	Yes	Yes	Yes
N	486	486	486	486	486
Adj R2	0.95	0.95	0.95	0.95	0.96

The table reports OLS estimates for the impact of employment density in on average productivity. The unit of observations are provinces. Regressions are weighted by total employment in year. Standard errors are clustered at province-level. Data source: SGK, Turkstat, WorldClim, Turkish State Meteorological Service

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Cem Özgüzel

June 11, 2020

A Background on Turkish context

A.1 Historical perspective on regional imbalances

Regional imbalances in Turkey go back to late Ottoman Empire when the geographical location of Western Anatolia, especially the major ports of export 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 shifted to Central Anatolia where the capital, Ankara, was established.

In the early Republic Period (1923-1950) nation-wide industrialization and urbanization policies were developed to promote the planning and development of settlement areas. These policies aimed at developing industries in major cities like Istanbul, Izmir, Ankara or Adana, but also some small Anatolian cities. Given that the State was a majority owner of commercial activity, these plans generated significant spatial transformation in the economic activities and population.

Starting in 1963, addressing regional inequalities became an official priority of the State. The State Planning Organisation (*Devlet Planlama Teşkilatı*) was given the specific responsibility of reducing regional disparities through multi-annual planning that foresaw the public investments.⁴³ State manufacturing investments, public enterprises and transport investments aimed to expand development in the poorer parts of the country, especially in Eastern and Southeastern regions.

In the late 1970s, the State's role as a technocratic agent of development shifted. During this period, the state began to recede from its interventionist mode to more of an enabler of the private sector (Boratav, 2017). In the third Development Plan (*Kalkınma Planı*, in Turkish) (1973-1977), Priority Provinces for Development (*Kalkınmada Öncelikli Yöreler*, in Turkish) all provinces of Eastern and Southeastern Anatolia were given priority in public investment. Going forward, all of the successive plans aimed to increase investment in these provinces both by increased public investment in the infrastructures and the offering of investment incentives to attract the private sector (such as tax break, lump-sum payments). In parallel, the State also ensured that all of the local administrative units received adequate

⁴³According to Gezici and Hewings (2004), despite focusing on the regional inequalities, these plans were not effectively implemented and thus did generate a significant reduction in inequalities. For a discussion on regional disparities in Turkey and government policies that aimed to tackle the issue see also Celebioglu and Dall'erba (2010).

financing capabilities.⁴⁴

After 1980, the new policies aimed at the development of the export base and favored the delocalization of the industrial activities from metropolitan cities to adjacent provinces of metropolitan regions. Regions located in the periphery of metropolitan cities started specializing in specific industries based on their comparative advantages in transportation network or natural resources. While the production moved to the periphery, Istanbul and other metropolitan areas, have increased its control on the private capital and management of the foreign trade (Gezici and Hewings, 2004). State continued also supporting private sector investment in less advanced regions and targeting public resources to carry out large infrastructure investments such as dams and new roads.

A.2 Urbanization

In 1960, Turkey still featured a largely agrarian economy with 31% of its population residing in urban areas.⁴⁵ While urbanization was steadily increasing during the 1950s-70s, it was during the 1980s that Turkey experienced a major surge of rural migrants to cities, causing a rapid expansion of informal areas in urban settlements.⁴⁶ During this period most of the rural migration were directed towards Turkey's three primary cities of Istanbul, Izmir, and the nation's capital of Ankara.

Between 2000 and 2010, Turkey's urban population has in a decade grown three times faster than its overall population. However, the growth during this period was driven by the country's secondary cities, which have experienced an increase of 53% of their population (World Bank, 2015). This growth was fueled by firms which were increasingly moving toward dynamic secondary cities, capturing economic spillovers from Turkey's large primary cities, while taking advantage of lower land rent values and labor costs. Turkey's principal cities, meanwhile, were diversifying their economies and focusing on innovation to remain competitive.

Overall since the 1960s, Turkey has structurally and demographically transformed from a predominantly agrarian economy to the globally competitive industrial economy it is today. During the country's most rapid period of urbanization, from 1960-2013, Turkey's industrial share of the economy increased from 17.6% to 27%, and the service sector dramatically rose from 26.4 to nearly 64%.

⁴⁴To ensure that local administration units did not suffer from lack of financing, the Turkish state founded the Bank of Provinces (*İller Bankası*, in Turkish) in 1933. The Bank was mandated to provide long-term funding that was necessary for local administrative units to finance the prepared municipal development plans and infrastructure projects. To this day, this Bank remains the principal source of financing for municipalities.

⁴⁵In terms of population, urban agglomerations were small in this period. Only one urban agglomeration (Istanbul) had more than 1 million, while 82% of the urban population lived in urban agglomerations with less than 500 thousand people. Between 1965 and 1980, the number of urban agglomerations with 1 to 5 million people increased from one to three. At that time, these large cities were home to about 39% of the country's population.

⁴⁶The mechanization of agriculture that began in the 1950s is one of the most significant transformations that took place in the modern Turkish economy (Pamuk, 2008) and triggered massive rural-urban migrations, which have profoundly affected the Turkish demography. Most of the migrants were originating from provinces with a large share of the agricultural sector in the eastern parts of the country moving to the more developed provinces in the west (Kirdar and Saracoglu, 2008). Many cities were unable to accommodate this growth, and the influx of migrants took place so quickly that these informal settlements became known as *gecekondu*, literally "houses erected overnight". States permissive policy toward rural migration and gecekondu encouraged migrants to flock to cities.

B Controls

In order to address the endogeneity bias due to the omitted variables, I add a number of variables drawn from the economic geography literature.

Land Area The productivity can be impacted by the local density but also by the size of employment. If gains from agglomerations outweigh the costs that are associated, both the density and the size of the local economy can have a positive impact on local productivity. To capture both effects, I add the province surface area, $area_p$. The impact of density, holding land area constant, reflects the gains from an increase in the number of workers in the province. It captures all the gains associated with a thicker labor market. The land area, holding density constant reflects the gains from increasing the spatial extent of the area (i.e., province). A larger area is likely to have more non-market interactions among agents than a smaller area as it is more populated. The data on the provincial land area come from Turkstat. It covers the real surface within the province borders, excluding lakes, in kilometre square (Km^2).

Market Potential Agglomeration effects may be operating beyond the borders of the unit of analysis. For instance proximity to larger outlets can be a source of profitability, as better market access allows firms to export more and at a lower cost (Krugman, 1980). In other words, better market access implies a stronger demand for the output of local firms and this increases the value of the marginal product of their labor and can thus be expected to lead to an increase in local wages. Furthermore, better access to outside markets makes local firms more productive by allowing them to access a broader variety of goods at a cheaper price, which again should lead to higher wages (Krugman and Venables, 2006).⁴⁷

The most common approach in estimating the market potential is to construct a measure following Harris (1954):

$$MP_{pt} = \sum_{i \neq p} \frac{den_{pt}}{dist_{ip}}$$

where MP_{pt} is the market-potential of province p in time t , which is equal to the sum of densities in the surrounding areas, divided by the distance between each market (i) and the province (p). This measure assumes that trade and communications costs are proportional to the inverse of distance. Furthermore, the own area is excluded (i.e., $i \neq p$) to avoid multicollinearity and also to identify separately the effects of the own location size (i.e. density) and external demand (i.e. market potential).

I calculate this variable by using employment density of the provinces (i.e. den_{pt})⁴⁸, divided by interprovincial bilateral distances provided by the General Directorate of Highways (*Devlet Karayolları*

⁴⁷See Fally et al. (2010) for the importance of market access in wages in Brazil, and Hering and Poncet (2010) in China.

⁴⁸Instead of density, the sum of salaries can also be used to capture the size of the local market. Results are robust to the use of either of the measures.

Müdürlüğü, in Turkish).⁴⁹ These distances measure the shortest route using real distances by road between the provincial centroids.

Beyond the domestic demand coming from locations within the country, labor productivity may also be affected by differences in foreign market access. Being close to trading ports or borders can lower the trade costs for areas, allowing the firms to access international markets and benefit from larger demand and externalities that may improve their productivity. In order to capture the foreign market potential, I use a similar measure:

$$FMP_{pt} = \sum_c \frac{GDP_{ct}}{dist_{cp}}$$

where the foreign market-potential of a province, FMP_{pt} is measured as the sum of GDPs of all the trading partners⁵⁰, divided by the distance between the capital of that country and the centroid of the province. The annual GDP data come from the World Bank. I constructed the distance measure by combining data from multiple sources. I obtained bilateral distance between Turkey and other countries from the *Gravity Database* provided by Centre d'Etudes Prospectives et d'Informations Internationales (CEPII) (Head et al., 2010). Using Googlemaps, I calculated the real road distance between the province centroid and the closest international port or border crossing through which international trade of that province is likely to pass. Finally, I combined the domestic distance and the international distance to obtain the bilateral distance between a province and a trading partner.

Specialisation Industrial employment can generate productivity gains both when it is higher because total employment at the location is higher, and when the share of the industry is higher for given employment at the location. While the former would be contributing to the urbanization externalities, thus benefiting all the sectors, the latter would provide additional productivity gains that are specific to the industry. To decompose these effects, I measure the employment share of the industry within the local economy in the following way:

$$spe_{pst} = \frac{emp_{pst}}{emp_{pt}}$$

where the spe_{pst} measures the specialisation of the industry s , in province p at time t , as the share of employment of that industry (emp_{pst}) in the total employment in the location (emp_{pt}). This measure should be included along with the total employment (or employment density) for a precise interpretation (Combes, 2000). Both of these variables are expected to have a positive impact, when there are urbanization and localization economies.

⁴⁹Using road distance is especially relevant in Turkey as the road transport is the primary mode of freight transportation in Turkey and accounts for 90% of domestic freight (by tonne-km) and passenger traffic (Cosar and Demir, 2016).

⁵⁰An alternative to the GDP is to use the value of the total trade between Turkey and the partner countries. My results are robust to using either of the two measures, and can be provided if requested.

Diversity The industrial diversity could also increase the local productivity by facilitating diffusion of knowledge within and between industries. This intuition made popular by Jane Jacobs (Jacobs, 1969) was formalized by Duranton and Puga (2001). Although various measures of diversity have been proposed, it is common to use the inverse of a Herfindahl index constructed from the shares of industries within local employment:

$$div_{pt} = \left[\left(\frac{emp_{pst}}{emp_{pt}} \right)^2 \right]^{-1}$$

where the div_{pt} is the degree of diversity in sectoral employment, in province p at time t , by the share of employment of that industry (emp_{pst}) in the total employment in the location (emp_{pt}).⁵¹ The measure captures the distribution of employment over all other industries and thus the urbanization economies that it generates.

Human capital externalities New Growth Theories emphasize the role of human capital as a determinant of productivity (Lucas, 1988). It is crucial to control for the skill-level in the location for two reasons. First, it is likely that locations with higher density also have greater share of skilled labor (Combes et al., 2008). If the differences in the skill levels are not controlled for, density will capture part of the gains due to compositional differences. Second, human capital accumulation generates benefits that go beyond the private gains to the worker and create positive externalities to those who are located in the area (Moretti, 1999). Introduction of control would allow for measuring the scale of these benefits.

The extent of human capital can be tested by adding a variable that captures the skills of the local labor force (Rauch, 1993; Moretti, 1999; Combes et al., 2011; Duranton, 2016). The linear approximation to a Cobb-Douglas production function requires adding the share of skilled labor as an explanatory variable:

$$HC_{pt} = \frac{Pop_{pt}^{hs}}{Pop_{pt}}$$

where HC_{pt} is a proxy for the share of skilled population in province p at year t , which is measured as the high-skilled population (Pop_{pt}^{hs}) with university degree or higher (ISCED4, ISCED5 and ISCED6) over the total population (Pop_{pt}).

The coefficient is not an elasticity, so its interpretation should be different. Including this variable allows capturing the effect of human capital on productivity. More precisely, it captures the positive impact of an increase in the share of high-skilled workers for a given level of density (or population size).

It is essential to underline that, given that I am using aggregate data, this variable captures both private

⁵¹Combes and Gobillon (2015) recommend exclusion of own industry when calculating this measure for a clearer interpretation especially if it is used along with specialisation. However, when the number of industries is large, it makes little difference to drop the own industry from computation since the correlation between two measures obtained with and without the own industry is large. In my case, given the large number of industries, I do not drop the own industry.

and social gains due to human capital.

Province level education data comes from Turkstat. The data is available at province level since 2008, includes the number of individuals by the highest diploma obtained (9 groups according to International Standard Classification of Education (ISCED, henceforth) classification and by sex.

Geographical Endowments Data on the distance from province centroid to the closest sea coast (*Coast*), length of seashores within the provincial boundaries (*Shore*), total length of rivers within the province (*River*) come from Sakalli (2019). See Sakalli (2019) for more details.

Cultural Amenities Data on amenities such as number of public libraries, cinema halls and public hospitals come from Turkstat. The data is provided annually for each province.

Weather All the data related to weather (i.e., mean average temperatures, average temperatures in January or June, average number of rainy days in a month and average number of sunny days in a month) is obtained from Turkish State Meteorological Service (*Devlet Meteoroloji Müdürlüğü*).

Provincial and village road networks I also collected data on the length of provincial road and village road networks from Turkstat. These numbers reflect the sum of road network in kilometers (Km).

C Summary Statistics

[Table C1 about here.]

[Table C2 about here.]

Table C1: Summary Statistics for Local Variables

Variable	Mean	sd	p10	p25	p50	P75	P90
LogDensity	2	1.1	0.6	1.1	1.7	2.4	3.2
Area (sq.km)	9,629	6,468.8	3,739.0	5,473.0	7,685.0	12,102.0	15,512.0
Diversity	37	14.3	21.1	27.2	34.9	44.5	55.3
Log Domestic Market Potential	1	0.5	0.3	0.5	0.9	1.2	1.6
Log Foreign Market Potential	23	0.1	23.1	23.2	23.3	23.4	23.5

Data source: Turkstat, SGK

Table C2: Summary Statistics for Individual Anaalysis

Variable	N	Mean	sd	Min	Max
Age	303,769	33.90	9.1	17.0	74.0
Female	303,769	0.23	0.4	0.0	1.0
Elementary School	303,769	0.02	0.1	0.0	1.0
High School	303,769	0.17	0.4	0.0	1.0
University	303,769	0.16	0.4	0.0	1.0

Data source: Household Labor Force Survey (Turkstat)

D Yearly results

The empirical analysis pools observations across 6 years to estimate the effect of employment density on local productivity. As a robustness test, I repeat the same exercise for each year separately. First, I present OLS results in Table D1, where the coefficients remain stable when the regression is repeated separately by year. Second, in Table D2 I present the 2SLS results where I instrument $\log Den_{pt}$ with population density in 1914 ($\log Den_{p1914}$). Overall the coefficients remain stable x which is expected given the stability of the relationship between density and productivity advantage of locations.

The yearly results also help address any concerns related to pooling of 6 years of observations together and instrumenting with time-invariant historical variables. As can be seen from the results, coefficients remain highly stable and statistically significant, while the standard errors are small.

[Table D1 about here.]

[Table D2 about here.]

Table D1: Yearly Regressions

	2008	2009	2010	2011	2012	2013
Density	0.0651 (0.004)***	0.0657 (0.004)***	0.0628 (0.005)***	0.0614 (0.005)***	0.0611 (0.004)***	0.0588 (0.006)***
N	81	81	81	81	81	81
Adj R2	0.85	0.85	0.82	0.82	0.84	0.78

The table reports OLS estimates for the impact of employment density in on average productivity. The unit of observations are provinces. Regressions are weighted by total employment in year. Standard errors are clustered at province-level. Data source: Turkstat, SGK

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D2: Yearly Regressions

	2008	2009	2010	2011	2012	2013
Density	0.0603 (0.004) ^{***}	0.0606 (0.005) ^{***}	0.0559 (0.005) ^{***}	0.0552 (0.005) ^{***}	0.0552 (0.005) ^{***}	0.0500 (0.006) ^{***}
N	78	78	78	78	78	78
r ²	0.84	0.85	0.81	0.81	0.83	0.76
Cragg-Donalds Stat	325.06	250.05	271.58	283.59	199.12	198.80

The table reports 2SLS estimates for the impact of employment density in on average productivity. The unit of observations are provinces. The number of observations is lower than OLS results due to differences in the number of provinces. See footnote 21 for details. Regressions are weighted by total employment in year. Data source: SGK, Karpat(1985)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

E Individual Analysis

E.1 First step individual regression

[Table E1 about here.]

Table E1: First Step Regressions, 2008-2013

	Log(Earnings)
Age	0.0607 (0.000)***
Age2	-0.0007 (0.000)***
Female	-0.1360 (0.001)***
Elementary School	-0.0129 (0.004)***
Middle School	0.0719 (0.005)***
High School	0.1593 (0.005)***
Vocational School	0.1796 (0.005)***
University	0.3921 (0.005)***
Specialisation	0.0279 (0.001)***
Occupation FE (39 groups)	Yes
Industry FE (87 groups)	Yes
RegionXYear FE (156 groups)	Yes
N	303769
Adj R2	0.60

The table reports OLS for the first step of the estimation. The dependent variable is log of earnings, regressed on individual characteristics, region-year fixed-effects (not reported) and industry fixed-effects (not reported). Table omits education group with no formal schooling ("No Schooling"). The unit of observations are individuals. Regressions are weighted by survey weights. Standard errors are clustered at region-level. Data source: Household Labor Force Survey (Turkstat)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

E.2 Robustness of the second step

In Section 5.1.2, I exploit the individual dimension of the labor Force Survey (LFS) to study the potential bias due to the sorting of workers across locations based on their observable and unobservable characteristics. In the second step of the estimation, I regress the region-year fixed-effects estimated in the first step, on the worker density of the region, which I compute also using the LFS for consistency. As a robustness, I repeat this estimation using density that is calculated using the SGK data. Table E1 presents these results.

[Table E1 about here.]

Table E1: Second Stage Regressions, 2008-2013

	No Controls		Individual Controls	
	OLS	2SLS	OLS	2SLS
Density	0.0580 (0.003)***	0.0527 (0.007)***	0.0507 (0.002)***	0.0483 (0.004)***
N	156	156	156	156
KP F-Stat		76.17		76.17

The table reports OLS and 2SLS estimates for the impact of employment density in on average productivity. The unit of observations are NUTS-2 Regions. Regressions are weighted by number of workers used in the first stage. Standard errors are clustered at region-level. Data source: Household Labor Force Survey (Turkstat), SGK, Kapat(1985)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

F Regions and provinces of Turkey

[Figure F1 about here.]

Figure F1: Regions and provinces



G Informal Employment

In developing economies, informal labor accounts for a substantial share of both urban and rural employment. According to Turkstat, the share of informal employment in the Turkish labor market stands high at 38.4% as of January 2012 (Turkstat, 2012). Moreover, Turkstat reports that the rate of informality to be 82.8% for the agricultural employment and 25.8% for the non-agricultural employment.

In this paper, I focus on the agglomeration economies based on formal employment. Given the prevalence of the informal sector, it is important to discuss the potential biases that could arise due to the omission of informal employment.

Given the objective of this paper, the exclusion of informal sector may matter for two reasons. First, it is important to understand the correlation between the share of informal employment and the density. If the distribution of informal employment as a share of the total employment is not homogenous across areas of different population sizes, this could create a bias in the estimated elasticity. Second, it can create omitted variable bias. I address these concerns in this section.

I use the Household Labor Force Survey as it allows distinguishing between workers who are affiliated to the social security vs. those who are not. Using this information I calculate the share of informal employment in the total employment in the region. The correlation between employment density (in levels) and share of informal employment is -0.22 (p-value<0.001). This suggests that areas with higher employment density have lower share of informal employment.

Furthermore, I estimate the impact of density on average wages including the workers employed in the informal sector. Specifically, I estimate the area-year fixed-effects (in the first step) and the density (in the second step) including workers who are not affiliated to the social security. Table G1 presents second step results. In Columns 1 and 2, I present the OLS and 2SLS results using only workers employed in the formal sector, for comparison.⁵² In Columns 3 and 4, I also include the workers who are in the informal sector.

[Table G1 about here.]

Results show that the inclusion of informal sector increases the estimated coefficient to 0.064. This suggests that i) the agglomeration economies matter for the informal sector, ii) these effects are stronger for the informal sector. The last result is in line with Atesagaoglu et al. (2017) who argue that exclusion of informal sector causes an underestimation of productivity in the Turkish context. In this sense, the elasticities estimated (using formal employment) in the paper present the lower bound estimates of the agglomeration economies. Although understanding the source of the difference in the elasticities is interesting, it goes beyond the scope of this paper.

⁵²I estimate the first step controlling for a full set of individual characteristics. These results correspond to the results in Column 4 in Table 5.

Table G1: Accounting for informal sector, 2008-2013

	OLS	2SLS	OLS	2SLS
Density (Formal)	0.0499 (0.003)***	0.0481 (0.005)***		
Density (Formal+Informal)			0.0638 (0.004)***	0.0631 (0.006)***
N	156	156	156	156
KP F-Stat		76.20		90.86

The table reports OLS and 2SLS estimates for the impact of employment density in on average productivity. The unit of observations are NUTS-2 Regions. Regressions are weighted by number of workers used in the first stage. Standard errors are clustered at region-level. Results in Columns 1 and 2 are obtained only using workers in the formal sector, while results in Columns 3 and 4 include also include workers with informal employment. Data source: Household Labor Force Survey (Turkstat), SGK, Karpat(1985)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

H First Stage of Multivariate Regression

Table H1 reports first stage results for the Table 6. Column titles refer to the endogenous variables that are instrumented. Titles above the results refer to the column numbers in Table 6.

[Table H1 about here.]

Table H1: First Stage Regressions, 2008-2013

	Column 3		Column 5		Column 7		Column 9		Column 11			
	LnDen	LnDen	LnDMP	LnDen	LnDMP	ISCED56	LnDen	LnDMP	ISCED56	LnDen	LnDMP	ISCED56
Density1914	1.2162 (0.188)***	1.2123 (0.187)***	0.0218 (0.019)	0.6511 (0.393)	0.0181 (0.069)	-0.0041 (0.003)	0.2469 (0.414)	-0.1479 (0.087)*	-0.0008 (0.002)	0.3689 (0.340)	-0.1193 (0.080)	-0.0010 (0.001)
DenGrowth	0.0031 (0.001)***	0.0032 (0.001)***	-0.0004 (0.000)**	0.0014 (0.001)	-0.0004 (0.000)**	0.0000 (0.000)	0.0006 (0.001)	-0.0007 (0.000)***	0.0000 (0.000)*	0.0015 (0.001)	-0.0004 (0.000)*	0.0000 (0.000)
DMP1945		-0.5908 (0.689)	1.7268 (0.146)***	0.2631 (0.738)	1.7325 (0.143)***	0.0063 (0.004)	-0.1578 (0.668)	1.6712 (0.124)***	0.0045 (0.004)	0.1803 (1.138)	1.6422 (0.216)***	0.0087 (0.005)
Enrolled Male 1927				0.0000 (0.000)*	0.0000 (0.000)	0.0000 (0.000)**	0.0000 (0.000)**	0.0000 (0.000)	0.0000 (0.000)*	0.0000 (0.000)	0.0000 (0.000)	0.0000 (0.000)*
Land Area							-0.4412 (0.349)	-0.1727 (0.050)***	0.0032 (0.002)*	-0.1200 (0.294)	-0.0892 (0.049)*	0.0024 (0.001)*
Diversity							0.9593 (0.456)**	0.2239 (0.083)***	0.0000 (0.001)	0.8798 (0.407)**	0.2310 (0.075)***	0.0001 (0.001)
Shores										0.0230 (0.056)	-0.0015 (0.010)	0.0006 (0.000)*
Coast										0.9531 (0.488)*	-0.0629 (0.115)	0.0052 (0.002)**
Climate										0.0361 (0.073)	0.0195 (0.015)	-0.0002 (0.000)
Rivers										-0.0007 (0.001)	-0.0007 (0.000)***	0.0000 (0.000)
N	468	468	468	468	468	468	468	468	468	468	468	468
Adj R2	0.81	0.81	0.76	0.83	0.76	0.58	0.85	0.80	0.63	0.87	0.83	0.68

The table reports first-stage results. Each column reports regression results where the variable on the column head is the dependent variable, regressed on various variables and controls. All the regressions include year fixed-effects and standard errors are clustered at province-level. Data source: Turkstat, Karpat(1985)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

