ECONOMIC RESEARCH FORUM
منتدى البحوث الاقتصادية
Working Paper Series

25th ERF ANNIVERSARY

AGGLOMERATION EFFECTS IN A DEVELOPING ECONOMY: EVIDENCE FROM TURKEY

Cem Özgüzel

Working Paper No. 1341

AGGLOMERATION EFFECTS IN A DEVELOPING ECONOMY: EVIDENCE FROM TURKEY

Cem Özgüzel[1]

Working Paper No. 1341

August 2019

I am grateful to Pierre-Philippe Combes for his guidance and support throughout this project. I thank Seyhun Orcan Sakallı for his generous help with the historical data, Güney Celbis, Hadi Esfahani, Martin Koning, Davide Luca, Sandra Poncet, Ariell Reshef and Ivan Savin, seminar participants at University of Paris 1, University of Strasbourg, 4th Workshop on Spatial Dimensions of the Labour Market (AMSE) 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.

Send correspondence to:
Cem Özgüzel
Université Paris 1 Panthéon-Sorbonne
cem.ozguzel@univ-paris1.fr

[1] Paris School of Economics. Université Paris 1 Panthéon-Sorbonne.

First published in 2019 by
The Economic Research Forum (ERF)
21 Al-Sad Al-Aaly Street
Dokki, Giza
Egypt
www.erf.org.eg

Abstract

Spatial inequalities in Turkey are a source of considerable policy concern. In this paper, I estimate agglomeration effects for Turkish provinces to shed light on the origins of spatial inequality in productivity and provide evidence from a developing country context which literature needs. I use social security data, an administrative dataset recently made available at the NUTS-3 level, for 81 provinces of Turkey for the period 2008-2013 and carry out a two-step estimation. I use 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 0.056-0.06, which is higher than in developed countries and around the levels observed in developing countries. Contrasting the evidence coming from developed countries, I find weak effects for sorting of workers across Turkish provinces based on observable characteristics.

Keywords: Local labour markets, spatial wage disparities, developing country, Turkey JEL
JEL Classifications: R12, R23, J31.

1 Introduction

In the past decade, the empirical literature on agglomeration economies provided robust evidence on the productivity gains associated with larger cities. Despite extensive evidence on these gains in developed countries, little is known about the impact of urbanization in the rest of the world. Addressing the knowledge gap regarding the urban economies in the developing world is important for two reasons. First, the majority of the world's urban population lives in countries that are far poorer than the advanced countries (e.g., US, Europe) where the evidence mainly comes from (Glaeser and Henderson (2017)). In addition to the importance of urban areas being drivers of economic growth in those countries, they also concern the lives of millions of people who reside and work in these places. Second, the models and stylized facts documented for cities in the developed countries may not apply entirely to the developing countries as characteristics and role of cities in the national economy may differ (Chauvin et al. (2017)). For instance, the rapid urbanization observed in the developing countries in the second half of the past century may have generated different benefits and costs compared to the western world where the urbanization rates have been relatively stable.

Turkey is an upper-middle-income developing country that has experienced fast urbanization and has a high rate of the urban population. Since the 1950s, the urban population in Turkey has increased dramatically due to massive rural-migration and high fertility rate. In 2017, 75% of the Turkish population lived in cities, making it a very highly urbanized country (World Bank).[1] In terms of GDP per capita, Turkey has the highest regional disparity among the OECD countries (OECD, 2018), while substantial spatial inequalities also exist in almost every metric (i.e., production, life quality, etc.). In this paper, I study the sources of spatial differences in productivity by focusing on agglomeration effects in Turkey. A better understanding of these effects is crucial for three reasons. First, productivity differences are an important driver of growth in Turkey, and thus, understanding which local factors make a given worker more productive is crucial. Second, understanding the determinant forces would make it possible to formulate policies to reduce regional differences. Third, with a population of 82 million, of which 75% percent living in cities makes agglomeration economies a relevant issue for a vast majority of the population. Despite the increasing urbanization and large spatial wage disparities, the determinants of these differences have not been studied yet. This study aims to fill this gap.

Spatial differences in wages (thus productivity) occur through three main channels.[2] First,

[1]Turkey went through rapid urbanization starting from the 1950s due to large rural-urban migration. While 34% of the population lived in cities in 1950, 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).

[2]Wages 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).

differences in wages across areas could be due to differences in skill composition of the workers. Workers with higher skills may sort into more denser areas, if for instance skill-intensive industries are not evenly distributed across areas. They could also sort into such areas if they value cultural amenities. Second, the differences in productivity could be due to local non-human endowments. For instance, workers in some areas may have a higher marginal product because of geographical features such as a favorable location (like a port or a bridge on a river), a climate more suited to economic activity, or some natural resources. Arguably, local endowments cannot be restricted to natural features and should also encompass factors of production such as public or private capital, local institutions, and technology. Third, interactions between workers or firms take place locally and lead to productivity gains. Following Marshall (1890), denser input–output linkages between buyers and suppliers, better matching of workers' skills with firms' needs in thicker labour markets, and technological externalities resulting from more intense direct interactions are frequently mentioned as the sources of these gains (see Duranton and Puga (2004) for a review). A key issue is to understand whether these benefits stem from the size of the overall market (urbanisation economies) or from geographic concentration at the industry level (localisation economies). Understanding of the spatial differences in wages thus requires consideration of all these explanations. Addressing all of these factors simuultaneoulsy would allow understanding and quantifying the relevance of each explanation. Understanding the magnitudes associated with each explanation is especially crucial for formulation of policy to address these inequalities.

This paper aims to take up on the challenge of explaining the determinants of wages disparities across Turkish provinces by applying the "uniformed approach" proposed by Combes et al. (2008). To do so, I use social security records, a novel administrative dataset that covers the complete universe of private sector workers that are affiliated to social security, working in all of the industry. This is the first paper to use this new dataset.

In order to provide estimates that are comparable to the literature, I follow the standard two-step approach similar to Combes et al. (2008).[3] This two-step approach allows me to distinguish gains due localisation economies (first-step) from those due to urbanization (second-step). Still, making causal interpretation of such forces requires addressing of two major identification issues: i) non-random spatial sorting of workers, ii) reverse causality from wages to city characteristics, if spatial wage differences are driving location choices of firms and workers.

The first concern requires accounting for spatial sorting of workers based on differences in their observable and unobservable characteristics. Addressing the sorting properly requires controlling of individual fixed effects through panel data. Such data however is unavailable in many developing countries, limiting the evidence from such contexts (Combes and Gobillon

[3] As argued by Melo in their meta-analysis, differences in unit of analysis, the specification of agglomeration economies and the choice of controls can give rise to significant differences in results reported in the literature. Using a standard empirical approach is thus crucial to provide comparable estimates. The two-step approach was first proposed by Combes et al. (2008), later to be followed by Martin et al. (2011); Bakens et al. (2013); De la Roca and Puga (2017); Combes et al. (2019) among many others.

(2015)). I circumvent this problem by complementing my analysis by using the Household Labor Force survey, which provides individual-level data. To address the endogeneity bias due to 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.

I find the elasticity of wages with respect to employment density to be around 0.056-0.06.[4] This means that doubling the employment density in an area increases the average wages by 3.8 - 4.2%. This elasticity is lower than one estimated for China (Combes et al., 2015; Chauvin et al., 2017) and India (Chauvin et al., 2017) and around those estimated for Brazil (Chauvin et al., 2017) and Colombia (Duranton, 2016). I also find a positive and strong effects for the domestic market potential. 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 important determinant of the productivity differences in Turkey. This means that 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. (2019).

I also do not find 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, 2019). This finding suggests that urbanisation 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, similar to the literature on developing countries. This pattern 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.

This paper contributes to the limited literature on the urban economics in developing countries by providing the first evidence from Turkey. It provides estimates from a highly urbanised, middle-income country and provide much needed evidence on the determinants of productivity in developing economies. I use a novel administrative data set, and use variety of panel data techniques and instrumental variables to deal with estimation concerns. I adopt a comprehensive and data intensive approach to provide estimates that are comparable with the rest of the literature. My findings corroborate earlier findings in the developing countries which 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 between the western cities and those in the developing part of the world.

The rest of the paper is organized as follows. In Section 2, I review the urban literature

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

and the evidence from developing countries to prepare the ground for my empirical strategy and contextualise my findings. I then present the Turkish context (Section 3) and present the data used (4). In Section 5, I present my empirical strategy and discuss the identification issues. Section 6 provides estimates on density and other determinants of productivity. Section 7 concludes the paper.

2 Literature Review

To base my empirical approach, I start by providing an extensive literature review on the estimation of agglomeration economies (Section 2.1) before discussing the magnitudes found in the literature (Section 2.2) and in the developing countries (Section 2.3).

2.1 Agglomeration Economies

The idea that larger cities enjoy a productive advantage dates back to Marshall (1890), who argued that larger markets benefit from more intensive input-out linkages, thicker local labor markets, and technological spillovers between firms which in return increase the average productivity. To explain differences in productivity, the literature offers three broad explanations.

The first explanation attests that larger cities have higher productivity as their market size facilitates sharing, learning or matching (Duranton and Puga, 2004). In a seminal paper, Ciccone and Hall (1996) tested this hypothesis by measuring the size of the local economy through the number of individuals per unit of land, that is density. Using aggregate data for American states, they studied the impact of the logarithm of density on the logarithm of workers' productivity, measured by nominal wage. The use of density was crucial for two reasons: first, it allowed overcoming concerns about the mismeasurement of the size of the local economy due to the heterogeneity in the spatial extent of the geographic units that are used in these studies.[5] Second, it allowed capturing the correlation between higher concentration of activity and the productivity, that is the combination of both agglomeration economies and dispersion forces.[6]

The second explanation puts forward that higher productivity is due to sorting of workers with higher abilities to larger cities. Glaeser and Maré (2001) was the first to test this idea by introducing individual fixed effects while studying the effect of density across US cities. The use of individual fixed effects was important as it allowed controlling for all the factors that can increase the individual's ability but remain constant over time. Since the individual fixed-effects can correlate flexibly with density and other local variables, it allows capturing

[5] The heterogeneity in the spatial extent of the geographic units and their shape (due to administrative borders) are problematic for two reasons. First, they make the comparison of estimates across studies difficult. Second, the size and the shape of the spatial units may influence the estimated elasticities due to "the modifiable areal unit problem". As shown by Briant et al. (2010), using density reduces concerns about both issues.

[6] It should be noted that although density captures the net effect of spatial concentration, it does not allow identification of specific channel through which it operates (Combes and Gobillon (2015)).

the effects of local characteristics net of composition effects due to sorting on the individual characteristics.[7]

The agglomeration effects identified in the early literature focused on the instantaneous effect of density on productivity. These gains which are specific to the location and thus *static,* augment the productivity of workers as long as they are located in the area. However, larger cities can also generate *dynamic* gains as they provide faster learning oppurtunities (Lucas, 1988; Glaeser, 1999). Futhermore, these gains can have long lasting effects if workers are able to transfer part of their productivity gains from agglomeration across locations. This would mean that a worker who moves from a larger city to a smaller one can bring part of his productivity and be more productiv than other individuals who have not worked in a large city.

Glaeser and Maré (2001) was the first to distinguish between the static and dynamic effects of agglomeration. Focusing on urban areas in Britain, D'Costa and Overman (2014) tested for both static and dynamic gains simulatenously. Using individual-level data for a large panel of British workers, they show that wage growth due to city size occurs in the first year that a worker moves to a city and this urban wage premium persists over time. In a more recent work, De la Roca and Puga (2017) proposed a dynamic framework that accounts for learning effects as a function of the time spent in different classes of city size. They argue that the knowledge that is acquired in larger cities is transferable over time and space. The degree of transferability of these dynamics gains, however, depends on the characteristics of locations.

2.2 Magnitudes for the effect of density on workers' productivity

The contributions in the literature differ in terms of context, data, empirical models and identification strategies which make the comparison of their results and conclusions difficult (Melo et al., 2009). Regardless of these difference, it has now been established that the local density of economic activities increases the productivity of firms and workers. This conclusion emerges from a large number of studies across different periods, contexts, and data.[8]

The earlier attempts to measure agglomeration economies used aggreagate data, where the logarithm of regional wage (or TFP) would be regressed on the logarithm of employment or population density. Two benchmark studies using aggregate data for the United States are Ciccone and Hall (1996) and Rosenthal and Strange (2008), who found an elasticity of productivity with respect to density, at around 0.04–0.05. This elasticity range implies that the doubling of density increases the average productivity by 3 to 4%. Similar elasticities were estimated for European regions (Ciccone, 2002; Brülhart and Mathys, 2008;Foster and Stehrer, 2009).

[7] This argument is further supported by Combes et al. (2008) who use an individual panel to estimate the effect of density on wages across all French cities. They show that not accounting for individual sorting (through fixed effects) leads to the overestimation of the coefficients of the local variables.

[8] Rosenthal and Strange (2004) and Combes and Gobillon (2015) provide an excellent overview of the literature. I only focus on recent articles that use richer datasets at the individual level that include workers' or firms' precise location, where the dependent variable is individual wages. See Combes and Gobillon (2015) for a review of the literature using TFP as a measure of productivity.

Glaeser and Maré (2001) were first to introduce individual fixed effects in estimation of the effect of density on wages using data from the American cities. Using a much larger panel for workers, Combes et al. (2008) estimate the effect of density on wages across all French cities using individual fixed effects and also taking into account aggregate endogeneity using a two-step estimation procedure involving instrumentation. They find an elasticity of wages with respect to density of around 0.030, which is half the elasticity that is obtained when individual unobserved heterogeneity is not taken into account. Using approaches accounting for individual heterogeneities through fixed effects or controls, similar elasticities are found for European economies (i.e. 0.025 for Spain by De la Roca and Puga (2017)); 0.01 for Italy by Mion and Naticchioni (2009); 0.016 for Britain by D'Costa and Overman (2014); 0.021 for Netherlands by Groot et al. (2014)).

In order to assess the magnitude of dynamic effects, De la Roca and Puga (2017) consider a quantity defined at the city level as the sum of the time-invariant city fixed effect and the effect of experience accumulated in the city for a worker who stayed there for 7 years (which is the average length of time for workers in their sample). The elasticity of this measure with respect to density (which captures both static and dynamic effects) is 0.049, which is almost twice as large as the estimated elasticity of city fixed effect of 0.025 suggesting importannt dynamic gains.

The presence of reverse causality between productivity and agglomeration has been addressed through various instrumentation techniques. The most common strategy is to use long-lagged values of population or population density to instrument present values of agglomeration economies (Ciccone and Hall 1996; Rice et al. 2006; Combes et al. 2008; Mion and Naticchioni 2009; De la Roca and Puga 2017) and geographical instruments (Ciccone, 2002; Rosenthal and Strange, 2008; Combes et al., 2008). The motivation for the choice of these instruments is that both past levels of urban size and external geologic variables (e.g., soil composition, depth to rock, water capacity, soil erodibility, and seismic and landslide hazard) are correlated with current levels of urban size but not with current levels of productivity (Combes et al., 2010). Correcting for aggregate endogeneity is has a small effect; sometimes decreasing the estimated elasticities by 10–20%, sometimes leaving the estimates unaffected or even increasing them slightly (Combes and Gobillon, 2015).

2.3 Evidence from developing countries

The literature from develeoping countries is recent and scant (Glaeser and Henderson, 2017). Furthermore, the lack of individual panel data in such countries makes it impossible to take into account unobserved individual heterogeneity. Differences between individuals have been taken into account through individual explanatory variables such as qualification, gender, age, and sometimes occupation or the type of firm where the individual is employed (Duranton, 2016; Chauvin et al., 2017; Combes et al., 2019).

Using individual data and standard instrumentation strategy, the effect of density on individual wages is found to be 0.05 in Colombia (Duranton (2016)), 0.09–0.12 in India (Chauvin et al., 2017) and 0.10-0.12 in China (Combes et al., 2015). Other measures of productivity have also been used in studies at the aggregate level, such as value added per worker in Korea (Henderson et al., 2001), establishment-level output per worker in Korea (Lee et al. (2010)), or output per worker in China (Au and Henderson, 2006), firm productivity in India (Lall et al., 2004) or in Chile (Saito and Gopinath, 2009).

The impact of market size on wages has been studied for China (Au and Henderson, 2006; Combes et al., 2013), India (Lall et al., 2004), and Colombia (Duranton, 2016).[9] Overall, theses studies find that market size has a larger effect than in developed countries.

Finally, some articles have studied local determinants of agglomeration economies other than market size such as industrial specialisation (Henderson et al., 2001 in Korea; López and Südekum (2009) in Chile) and industrial diversity (Saito and Gopinath (2009) in Chile).

3 The Turkish context

Turkey has a population of about 81 million over an area of 783 thousand square kilometers (Turkstat, 2019). While 74 percent of the country's population live in cities, some 92 percent of Turkey's gross value added is produced in cities (World Bank, 2015).

With a Gross Domestic Product (GDP) per capita of USD 10546 in 2017, Turkey is an upper-middle-income developing country (according to the World Bank classification). 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). These differences are multiplied even further 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 percent of the national GDP, adding its immediate surrounding area increases the share to 39 percent in 2017 which creates large imbalances across regions in terms of production and income (Turkstat, 2017).[10]

Regional Imbalances Regional imbalances in Turkey go back to the 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

[10] The surrounding includes Kırklareli, Tekirdağ, Kocaeli, Yalova, Sakarya and Bursa provinces.
9 Similarly, the impact of market access on individual wages is also estimated in Brazil (Fally et al., 2010) and China (Hering and Poncet, 2010).

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 developped to promote the planning and development of settlement areas, developing industries in major cities like Istanbul, Izmir, Ankara or Adana, but also in selected small Anatolian towns. 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, adressing regional inequalities became an official priority of the State. The State Planning Organisation (*Devlet Planlama Teşkilatı*) was given the specific responsability of reducing regional disparities through multi-annual planning that foresaw the public investments.[11] 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 70s, 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 financing capabilities.[12]

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.

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

[12] To 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 today, this Bank remains the principal source of financing for municipalities.

Urbanization In 1960, Turkey still featured a largely agrarian economy with 31 percent of its population residing in urban areas.[13] 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 rapid expansion of informal areas in urban settlements.[14] 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 percent to 27 percent, and the service sector dramatically rose from 26.4 to nearly 64 percent.

3.1 Literature on regional differences

Turkish regional imbalances has been the focus of a number of papers over the years. While some 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; Luca, 2016).

Although there has been a considerable number of studies dealing with the determinants of overall productivity(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 change of the private and public sectors in the Turkish manufacturing industry in eighteen provinces from 1990-1998 or Temel et al. (1999) who use gross provincial product

[13] In terms of population, urban agglomerations weree small in this period. Only one urban agglomeration (Istanbul) had more than 1 million, while 82 percent of the urban population lived in urban agglomerations with less than 500,000 people. Between 1965 and 1980, the number of urban agglomerations with 1 to 5 million people grew to 3. At that time, these large cities were home to about 39 percent of the country's population.

[14] The 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 east 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 gecekondus encouraged migrants to flock to cities.

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 have also 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 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 endogeneity of local determinants of agglomeration economies.

This paper is the first evidence from Turkey focusing on the spatial productivity differences 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 the 2000s.

4 Data and Sample

This section presents different sources of data that are used to estimate the determinants of spatial differences in wages.

4.1 Data

4.1.1 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 sector.[15]

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, total number of days worked, 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

[15] The data covers all employment with compulsory insurance in the private sector under Article 4-1/a of Act 5510. For the year 2013, this corresponds to 12,5 million individuals. 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, nor those who are employed in the public sector (2.8 million) under Article 4-1/c of Act 5510.

(659 sectors) for the period 2007-2013. The data is further disaggregated by job contract-type (temporary vs. permanent), by sex (male vs. female).

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.[16] This dataset makes it possible, for the first time, to focus on productivity differences at a geographically more disaggregated level. This is also the first paper to use this dataset.

4.1.2 Household Labor Force Survey

I complement the main analysis by using individual level data obtained from the Household Labor Force Survey (*Hanehalkı İşgücü Anketi*, LFS henceforth) prepared by the Turkish Statistical Institute (*Türkiye İstatistik Kurumu*, Turkstat henceforth). The main objective of the labour force survey is to obtain information on the structure of the labour force in the country. The national labor force statistics, for instance, are produced based on the LFS surveys. It is representative of the total population in Turkey. It includes annual information on economic activity, occupation, status in employment and hours worked for employed persons (both formal and informal sector); unemployment, education and much more.

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

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

4.1.3 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. 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).[17] I complement this data by digitizing published census reports for the

[16] Two 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.

[17] Between 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).

period 1928, 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 coming from the Maddison Project Database (Bolt and van Zanden, 2014).

4.1.4 Controls

Land Area The data on the provincial land area come from Turkstat. It covers the real surface within the province borders, excluding lakes, in kilometer square.

Education Province level education data comes from Turkstat. The data is available at province level since 2008, includes the number of inviduals by the highest diploma obtained (9 groups according to International Standard Classification of Education (ISCED, henceforth) classification and by sex. I use this data to calculate the human capital measure which I explain in detail in Section 5.2.1.

Road Lengths I use real road lenghts to measure distance between provinces. The data comes from the General Directorate of Highways (*Devlet Karayolları Müdürlüğü*, in Turkish). These distances measure the shortest route using real distances by road between the provincial centroids. I use this data to calculate market potentials. I also collected data on the lenght of provincial road and village road networks from Turkstat. These numbers reflect the sum of road network in kilometers (Km).

Foreign Market Potential I calculate the Foreign Market Potential using the GDP of the trading partners and bilateral distance between Turkey and other countries using the *Gravity database* provided by Centre d'Etudes Prospectives et d'Informations Internationales (CEPII) (Head et al., 2010).

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 Meteroloji Müdürlüğü*).

4.2 Sample

For my analysis I use data on workers from private sector, holding temporary and permanent job contracts, for both females and males. Thus my sample covers all the workers in the formal sector. I check the sensitivity of my results for subgroups as a robustness.

I focus on the period 2008-2013. I exclude 2007 as data on skill-levels is not available for that year. This leaves me with a final sample of 168 904 industry-province-year observations which I use to estimate the province-year fixed effects in the first step.

5 Empirical Strategy

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

5.1 Econometric Equation

To evaluate the impact of agglomeration economies on the productivity, the following specification can be used:

$$logw_{pst} = \alpha + \beta logDen_{pt} + \gamma_s + \gamma_t + \varepsilon_{pst} \tag{1}$$

where w_{pst} is the average daily wage in province p, sector s at time t,[18] den_{pt} is the total number of employees (or population[19]) in province p at time t (emp_{pt}) divided by land area ($Area_p$).[20]

Since 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

[18] I calculate the daily wage by dividing the average monthly salary by the number of days worked.

[19] Depending 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.

[20] Formally, this corresponds to: $Den_{pt} = emp_{pt}/area_p$.

are used. Moreover, the use of density adresses 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 litearture.[21] 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 and negative.

The panel structure of my 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), and time fixed-effects (γ_t) to absorb 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 deflator (Combes and Gobillon, 2015). Moreover, nominal 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).[22] Finaly ε_{pst} is the error term representing unexplained productivity.

5.2 Estimation Issues

There are four sources of bias in the estimation of Equation 1. First, some other province characteristics may be correlated with density and wages. In that case employment density may capture the impact of omitted variables that may be determining the productivity in the province. Second, workers may prefer to locate in cities where wages are higher. This creates a reverse causality problem. Third concern is related to the computation of standard errors. Fourth, workers may sort across locations based on their ability. In this section, I address these issues by proposing multivariate estimations to take multiple variables into account, and then to instrument some of them.

5.2.1 Estimation issue 1: Omitted Variables

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 the local productivity. In order to capture both effects, I add the province surface area, $area_p$. The impact of density,

[21] For a discussion on why the use of density reduces the mismeasurement of the size of the local economy, see Briant et al. (2010)

[22] 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.

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 surface area of each province in Km^2 come from Turkstat.

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).[23]

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 commmunications costs are proportional to the inverse of distance. Furthermore, the own area is excluded (i.e., $i \neq p$) to avoid multi-collinearity and also to identify sepearately 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})[24], divided by interprovincial bilateral distances provided by the General Directorate of Highways (*Devlet Karayolları Müdürlüğü*).[25] 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

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

[24] Instead of density, the sum of salaries can also be used to capture the size of the local market. My results are robust to the use of either of the measures.

[25] Using 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).

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 partners[26], 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 used the bilateral distance between Turkey and other countries provided by CEPII to measure the international distance. 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.

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

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

[26] An 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.

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}).[27] 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 and social gains due to human capital.

Amenities Local amenities determine both local density and wages. Productive amenities such as airports, transport infrastructures, and universities increase productivity and attract workers, which makes the density increase.[28] These endowments can raise wages through various channels by lower exporting costs, cheaper supplies, or higher productivity. In that case, a positive bias in the estimated coefficient of density is also expected. Using a complete

[27] Combes and Gobillon (2015) recommend exclusion of own industry when calculating this measure for a clearer interpreatation especially if it is used along with specialisation. However, when the number of industries is large, it makes little difference to drop the own indsutry 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.

[28] Arguably, local endowments cannot be restricted to natural features and should also encompass factors of production such as public or private capital, local institutions, and technology.

set of productive endowments would, however, raise serious endogeneity concerns (Combes and Gobillon, 2015).

As put forth by Roback (1982), consumption amenities such as cultural heritage, social life, or climate-related amenities can increase the attractiveness of some locations for workers and thus make density higher. Such amenities do not have any direct effect on productivity, but the increase in housing demand they induce makes land more expensive. As a result, local firms use less land relative to labor, and this decreases labor productivity when land and labor are imperfect substitutes. This causes a negative bias in the estimated coefficient of density since density is positively correlated with missing variables that decrease productivity. To avoid the bias, in my benchmark analysis I control for endowments that are exogenous such as distance to sea coast, length of rivers, and lenght of seashores and average temperatures.

In the final part of the paper, I extend further the number of controls and add climate-related amenities, which have the advantage of being exogenous to the local economy but also being well measured (Chauvin et al., 2017). Specifically, I include variables that account for average temperatures in January and its difference from the ideal temperature of 21.1 Celsius. I also add the average yearly number of sunny days and annual rain volume. All of the data is obtained from the Turkish State Meteorological Service (*Devlet Meteroloji Müdürlüğü*).

Despite endogeneity concerns, I also provide additional results with other types of productive (e.g., road network) and non-productive amenities (e.g., cinema halls, hospitals, libraries) as it is informative. I construct all of these controls using data provided by Turkstat.

5.2.2 Estimation Issue 2: Circular Causality

The second estimation issue is that some local characteristics are likely to be endogenous to local wages. For instance, employment areas receiving a positive technology shock may attract migrants. This leads to a positive correlation between the residuals and the density of employment. In this particular case, reverse-causality is going to bias the estimates upwards. Other regressors such as market potential or human capital are 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, 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 population growth. The intuition is that current productivity shocks are not correlated with the employment structure from decades before the date of observation.

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. In addition, considerable population shifts took place between 1914-1924, 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 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 Section 3, 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).

I construct past domestic market potential using 1935 and 1945 population numbers, and past foreign market potential using the GDP levels in 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, since the population is taken in logarithmic form, using a multiplicity of census dates is equivalent to instrumenting by past levels and long-run historical growth rates. Second, having multiple instruments allows me to the instrument not only for employment density but also for the market potential, diversity, and even land area. I can also conduct exogeneity and 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 characteristics 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.[29]

Across the tables, I use two groups of historical variables to instrument for density. The first group consists of population density in 1914, and the population growth rate between 1914 and 1927. 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 choose to be parsimonious and report estimations for different groups of workers using the same sets of instruments to allow for reliable comparisons.

[29]For a similar argument see Combes et al. (2015, 2019).

5.2.3 Estimation Issue 3: Accounting for local shocks

Equation 1 can be estimated, including the controls, and the IV strategy explained before. 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. I estimate the following equations:

$$logw_{pst} = \alpha + \delta logSpe_{pst} + \gamma_s + \gamma_{pt} + \varepsilon_{pst} \tag{2}$$

$$\gamma_{pt} = \nu + \beta_1 logDen_{pt} + \theta X_{pt} + \phi Z_p + \gamma_t + \varepsilon_{pt} \tag{3}$$

In the first step (Equation 2), I regress the log average daily wages $(logw_{pst})$ in province p, sector s at time t on $logSpe_{pst}$ which captures the effect of specialisation in a given sector on productivity, sector fixed-effect (γ_s) and province-year fixed-effect (γ_{pt}). The sector fixed-effects (γ_s) capture any sector-specific differences in the productivity that is irrespective of time, while location-year fixed effect can be interpreted as local wage indices after controlling for observed and unobserved industry effects. ε_{pst} is the error term.

The province-year fixed effects estimated in the first step are then used as the dependent variable in the second step (Equation 3) and regressed on local characteristics that impact the productivity levels. To account for the local structure I use density $(logDen_{pt})$ but also X_{pt}, which includes time-varying controls (e.g. market potentials, diversity, human capital, road network, and more.) and Z_p, which includes the time-invariant controls (e.g., land area, temperatures, geographic controls, etc.). γ_t is the year-fixed effect which, as discussed before, takes care of any shock that impacts the productivity levels across the whole country, and also correct for the use of nominal wages which are not deflated.

This method is preferable for two reasons. First, as explained before, doing a two-stage 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 $(logSpe_{pst})$.

Second, this procedure makes it possible to sepearately 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.

As a robustness check, in Appendix Section C, I run a single-stage estimation and find very similar results.

5.2.4 Estimation Issue 4: Sorting by ability

The final identification problem is the possible correlation between density and worker characteristics. If the workers' spatial distribution depends on their abilities, then the local productivity would also be affected by the differences in the composition of the 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 also due to differences in the composition of the workers. These differences can be due to observed and unobserved ability.

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. Differently than these developed countries, Combes et al. (2015) find very weak sorting 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 for developing countries. An alternative solution is to control for an extensive set of individual characterists to take care of differences in observable skills (Duranton 2016; Chauvin et al. 2017; Combes et al. 2019). If workers 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. Given that I am using data aggregated at industry-province level, the estimated coefficients would be inflated if there is sorting bias. It is thus important to see whether such bias exists in Turkey, and if so, measure its size. To measure the potential bias due to sorting, I use Household Labor Survey and exploit its individual-level dimension.

6 Main Results

In this section, I start by estimating the elasticity of wages to density by addressing gradually each identification concern discussed previously (Section 6.1). Then I present results of a multivariate framework accounting for local characteristics that determine the local productivity (Section 6.2). I end the section with estimations that include infrastructure and amenities (Section 6.3).

6.1 Density

I start with estimating the effect of density on average productivity. As discussed earlier, an unbiased estimate can be obtained if the coefficient is not suffering from identification issues explained previously. I explore the elasticity of density through a simple framework where these concerns are addressed separately, before moving to more complex models with multiple controls and instruments.

6.1.1 Omitted Variable Bias

I apply two-stage regression where in the first-stage I estimate province-fixed effects which are used as the dependent variable in the second stage.[30] Table 1 presents OLS results for the second stage 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 coefficients (Solon et al. (2015)). Standard errors are clustered at province-level.

Column 1 of Table 1 shows that density has an elasticity of 0.06. This means that doubling the worker density increases the average productivity by 4%.[31] If density of Iğdır (2.96, P25) were to increase to density of Mersin (11.3, P75), its productivity would increase by 14.7%.[32]

In the following columns, I address the concern due to omitted variable bias by adding a number of controls that are standard in the literature.[33] 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 liter-ature. The positive coefficient suggests that, for a given density, a 1 percent 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

[30]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 percent increase in the specialisation of the local industry increases the average wages by 0.065 percent.

[31]$2^{0.06} - 1 \approx 4\%$

[32]This elasticity is the result of pooling observations across 6 years. However, as can be seen in Appendix Table B1, the coefficient remains stable when the regression is repeated separately by year.

[33]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.

significant. The elasticity of density drops to 0.051, suggesting that provinces with higher employment density arealso more diverse. 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.

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. 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 percent, 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.[34] 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.

6.1.2 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 important 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 identificaiton strategy consists in implementing the 2-stage-least-squares (2SLS) estimation outlined in Section 5. 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 historical instruments that are based on historical census data from 1914, 1927, 1935 and 1945. Specifically, I instrument $logDen_{pt}$ with population density in 1914 ($logDen_{p1914}$), population density in 1927 ($logDen_{p1927}$), employment density in 1935 ($logDen_{p1935}$), employment density in 1945 ($logDen_{p1945}$), 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 2008-2013. Each observation in the figure corresponds to a province.

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

The figure shows that density in 1914 is a good predictor of the density today. As shown in the first-stage regressions presented in Table 2, due to the strong inertia of the urban hierarchy in Turkey, population densities at the beginning of the 20th century are good predicter for the employment density in 2008-2013 period. However, given the significant changes that took place between two periods, it is unlikely that they are correlated to labor productivity.

I formally test the validity of each instrument by reporting the first-stage regression using the following equation:

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

Table 3 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.

Each column presents the coefficients coming 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 2008-2013.

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.[35] As expected the F-statistics gets larger when I use densities from more recent periods.

Table 3 presents the results for the estimation of Equation 3.[36] As explained previously, the parameter β_1 corresponds to the effect of employment density on the productivity levels of the provinces.

In the first column, I report the elasticity obtained through OLS estimation. In the following columns, I present 2SLS results where the variable 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.

[35]The 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.

[36]Appendix Table A1 reports the summary statistics for the second-stage estimation.

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

Table 1 shows that 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 specification. 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.

Compared to estimates in other developing countries, the elasticity of density is slightly higher than 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., 2013). Given the level of urbanization in Turkey, this elasticity fits precisely where it would be expected. However, these last papers use individual-level data, which allows them to net out the endogeneity due to the possible sorting of the higher ability individuals into denser areas. Thus their elasticities present gains purely due to agglomeration economies generated by the higher densities. As I am using aggregate data, the elasticity that I estimate however includes the positive bias due to individuals sorting. This means that these elasticities can be biased up to 100% as shown, in the context of France, by Combes et al. (2008) although 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., 2015, 2017). I explore this issue further in the next section.

6.1.3 Sorting By Ability

As explained in Section 5.2.4, the paramaters estimated in the second-step can be biased if workers with higher abilities sort into denser areas. 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. 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 sex. It is sampled to be reflective of the Turkish population but also 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 aggreagated at NUTS-2 regional level. This means that the locations of individuals can be identified only at one of the 26 regions.[37] Use of smaller scale sizes are better for capturing benefits of interactions that decay with distance (Combes and Gobillon, 2015). Still, using larger local units may 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 individiual fixed effects in my estimation similar to Combes et al. (2008). In order to examine the existence of selection, I follow Combes et al. (2015) and carry out two tests[38]: first, I estimate the first step regressions by successively including different sets of explanatory variables, two, I compare the second step regression using fixed-year effects which were estimated in a first-step including in individual characteristics vs. those which were not.

First, I estimate the first step estimation including different sets of explanatory variables (location effects, individual characteristics, and firm characteristics) to understand their relative contributon to the log of the monthly wage. Table 4 reports the adjusted R^2 of each regression.[39]

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 dummies and specialisation together 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.

These results reveal that these three sets of effects are fairly orthogonal. Region effects and individual characteristics together explain 44% of the wage disparities when 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. 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. (2019) for China.

[37] 26 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.

[38] This approach is also used in Combes et al. (2019) for China, and Colombia in Duranton (2016).

[39] Full estimation results are available if requested.

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.[40] 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 use the waves for the Household Labor Force survey for the period 2008-2013, and create a sample that matches the SGK data. Similar 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.[41] 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.[42] This leaves me with 412 137 individual observations over the 6 year period.

I use the two-step procedure with individual level data similar to Combes et al. (2008). The procedure consists in estimating the followinig specification:

$$logw_{irst} = \alpha + \delta logSpe_{rst} + \phi X_{it} + \gamma_s + \gamma_{rt} + \varepsilon_{irst} \tag{4}$$

$$\gamma_{rt} = \nu + \beta_1 logDen_{rt} + \gamma_t + \varepsilon_{rt} \tag{5}$$

The first-step estimation of equation 4 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 X_{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 ($logDen_{rt}$) and time-fixed effects, γ_t. I measure density using the survey data for consistency.[43]

[40] See Appendix Table D1 for the first-step results.

[41] There are a few reasons I apply this 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 E, I provide estimates including also workers in the informal sector.

[42] Although not reported, I 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 those who are not affiliated to the social security increases the elasticity to 0.063-0.07. 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. Inclusion of self-employed does not change elasticities. Results are available if requested.

[43] In Appendix Table D1, I test the robustness of my estimate by measuring employment density using the SGK employment data. Results are almost identical.

Table 5 reports OLS and 2SLS results for the second-step.[44] 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 stage is weighted with survey weights, the second stage is weighted with the number of workers used to estimate the region-year fixed effects in the first stage. 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. The difference in the estimated elasticities in Columns 2 and 4, suggest that excluding individual controls only inflates the elasticity by 8%. This small number shows that the workers in Turkey do not sort across locations based on their observables.

As mentioned earlier, the sorting can also be based on the unobservable characteristics of the individual. However, as argued by Combes et al. (2015), it is very unlikely for a sorting based on unobservables to take place while sorting on observables is so weak. Granted, a final conclusion on the issue can only be given following an analysis using panel data. Such data however is currently unavailable in Turkey.

Second, the similarity between these findings and those found in China (Combes et al., 2015, 2019) 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 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.

[44] First stage results are presented in the Appendix Section D.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 us to account for the individual characteristics (e.g. education, occupation) and estimate 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.

6.2 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 3, 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. (2019) 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 $logDen_{p,t}$ with population density in 1914 $(logDen_{p1914})$ and growth in population density between 1914 and 1927 $(DenGrowth_p)$; domestic market potential $(logDMP_{pt})$ with domestic market potential in 1945 $(logDMP_{p1945})$; and human capital (HC_{pt}) with number of enrolled male students in 1927 $(EnrolledMale_{p1927})$.[45]

In order to test the statistical relevance of these instruments, I report Cragg–Donald F-Statistic weak instrument test, Shea's partial R^2 which shows that my instruments explain a large share of the variation in the instrumented variables, once potential inter-correlations among instruments have been accounted for, and finally, the Hansen J-Statistic tests over-identifying restrictions.

Table 6 presents the 2SLS results.[46] For comparability, I start by presenting the OLS and 2SLS results where density is the only explanatory variable (Columns 1 and 2). In Column 3 and 4, I add domestic market potential, which I instrument with the domestic market poten-tial in 1945. Introduction of this additional variable does not change the elasticity of density. In Columns 5 and 6, 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

[45]As mentioned earlier, I have a larger set of possible instruments and I experimented with multiple combinations. Estimations using various combinations yielded largely consistent results with one another. I choose to be parsimonious, and report estimations using the same sets of instruments to allow for reliable comparisons. I also try to be restrictive about the number of instruments used and use just enough to carry out over-identification tests.

[46]Appendix F1 table reports the first-stage results.

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.[47]

In Columns 7 and 8, I further control for the land area and diversity of the local economic activity. Although the former can be considered exogenous to the density, the latter is correlated. As the expected land area is highly significant, and for a given density, an increase in the land area increases the average productivity. If the land size of a province doubles, the 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 9 and 10, 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*).[48] 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.[49] 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 (10) 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, Land Area* and *Diversity*) and amenities (*Shore, Coast, Climate, 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 0.064, 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. 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.

[47]The 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.

[48]I drop diversity as it is insignificant and generates unnecessary endogeneity.

[49]It 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.

(2008), but smaller than 0.13-0.22 found for China in Combes et al. (2019).

6.3 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 average wages are also determined by the infrastructure and such 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 reports the OLS results where I augment the model in Section 6.2 with additional controls. In all of the regressions I control for market potential, human capital, geographic endowments, land area and include time-fixed effects. In Appendix Table G1, I also report the same results without these controls.

Column 1 replicates the results from Table 6 for comparison. In Column 2, I account for the lenght of provincial road network (*Roads*) and village roads (*Village Roads*). The results in the column suggest that denser road network does not impact the wages.[50] 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.[51]It should also be noted that these coefficients capture the effects that remain when controlling for domestic market potential.

[50]It 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.

[51]For more on the issue, see Duranton (2016).

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.[52]

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.[53] 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.[54] 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 OLS results should be viewed as providing robustness checks for my 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. Although not presented, all of the amenities are correlated with density and other controls. 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, this would indicate the weak relationship between productivity (wages) and amenities in a developing country context.

7 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. In

[52]Although 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.

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

[54]This 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.

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 6 percent 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 literature, I find a positive effect on market access, which is much stronger than those found in developed countries.

I also find evidence for very weak sorting of workers across provinces based on their observed abilities. This finding contrasts with 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. Although this conclusion needs to be corroborated with further evidence coming from a broader range of developing countries, a better understanding of the sources of these differences are essential missing pieces in the literature, and thus remains high on my research agenda.

References

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.

Bakens, Jessie, Peter Mulder, and Peter Nijkamp. "Economic impacts of cultural diversity in the netherlands: Productivity, utility, and sorting*." *Journal of Regional Science* 53, 1: (2013) 8–36.

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.

Baslevent, 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." *The Economic History Review* 67, 3: (2014) n/a–n/a. `http://doi.wiley.com/10.1111/1468-0289.12032`.

Boratav, Korkut. *Türkiye Iktisat 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.

Brülhart, Marius, and Nicole A. Mathys. "Sectoral agglomeration economies in a panel of European regions." *Regional Science and Urban Economics* 38, 4: (2008) 348–362.

Celbis, Mehmet Güney, and Denis de Crombrugghe. “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. “Urbanisation and migration externalities in China.” .

———. “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* .

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, Gilles Duranton, Laurent Gobillon, and Sébastien Roux. *Estimating agglomeration economies with history, geology, and worker effects*, volume I. Cambridge, MA: NBER, 2010.

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.

Foster, Neil, and Robert Stehrer. “Sectoral productivity, density and agglomeration in the wider Europe.” *Spatial Economic Analysis* 4, 4: (2009) 427–446.

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. “Learning in Cities.” *Journal of Urban Economics* 46, 2: (1999) 254–277.

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.

Glaeser, Edward L., and David C. Maré. “Cities and Skills.” *Journal of Labor Economics* 19, 2: (2001) 316–342.

Groot, Stefan P.T., Henri L.F. de Groot, and Martijn J. Smit. "Regional wage differences in the netherlands: Micro evidence on agglomeration externalities." *Journal of Regional Science* 54, 3: (2014) 503–523.

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.

Henderson, Vernon, Todd Lee, and Yung Joon Lee. "Scale externalities in Korea." *Journal of Urban Economics* 49, 3: (2001) 479–504.

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* .

Lee, Bun Song, Soomyung Jang, and Sung Hyo Hong. "Marshall's scale economies and Jacobs' externality in Korea: The role of age, size and the legal form of organisation of establishments." *Urban Studies* 47, 14: (2010) 3131–3156.

López, Ricardo A., and Jens Südekum. "Vertical industry relations, spillovers, and productivity: Evidence from chilean plants." *Journal of Regional Science* 49, 4: (2009) 721–747.

Luca, Davide. “Votes and Regional Economic Growth: Evidence from Turkey.” *World Development* 78: (2016) 477–495.

Lucas, RE Robert E. “On the mechanics of economic development.” 22, 1: (1988) 3–42.

Marshall, Alfred. *Principles of Economics*. London: Macmillan, 1890.

Martin, Philippe, Thierry Mayer, and Florian Mayneris. “Spatial concentration and plant-level productivity in France.” *Journal of Urban Economics* 69, 2: (2011) 182–195. https://www.sciencedirect.com/science/article/pii/S0094119010000665.

Melo, Patricia C. PC, Daniel J. DJ Graham, and Robert B. RB Noland. “A meta-analysis of estimates of urban agglomeration economies.” 39, 3: (2009) 332–342.

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.

Mion, Giordano, and Paolo Naticchioni. “The spatial sorting and matching of skills and firms.” *Canadian Journal of Economics* 42, 1: (2009) 28–55.

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. *OECD Regions and Cities at a Glance 2018*. 2018.

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.

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. “Redding and Venables 2003 Economic Geography and International Inequality.” *Journal of international Economics* .

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.

Rice, Patricia, Anthony J. Venables, and Eleonora Patacchini. "Spatial determinants of productivity: Analysis for the regions of Great Britain." *Regional Science and Urban Economics* 36, 6: (2006) 727–752.

Roback, J. "Wages, rents, and the quality of life." *Journal of political economy* .

Rosenthal, Stuart S., and William C. Strange. "Evidence on the nature and sources of agglomeration economies." In *Handbook of Regional and Urban Economics*, Elsevier Inc., 2004, volume 4, 2119–2171.

———. "The attenuation of human capital spillovers." *Journal of Urban Economics* 64, 2: (2008) 373–389.

Saito, Hisamitsu, and Munisamy Gopinath. "Plants' self-selection, agglomeration economies and regional productivity in Chile." *Journal of Economic Geography* 9, 4: (2009) 539–558.

Sakalli, Seyhun Orcan. "Secularization and Religious Backlash : Evidence from Turkey." , 2014.

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.

World Bank. "Rise of the Anatolian Tigers." Technical Report 87180, The World Bank Group, Washington DC, 2015.

Figure 1: Spatial Differences

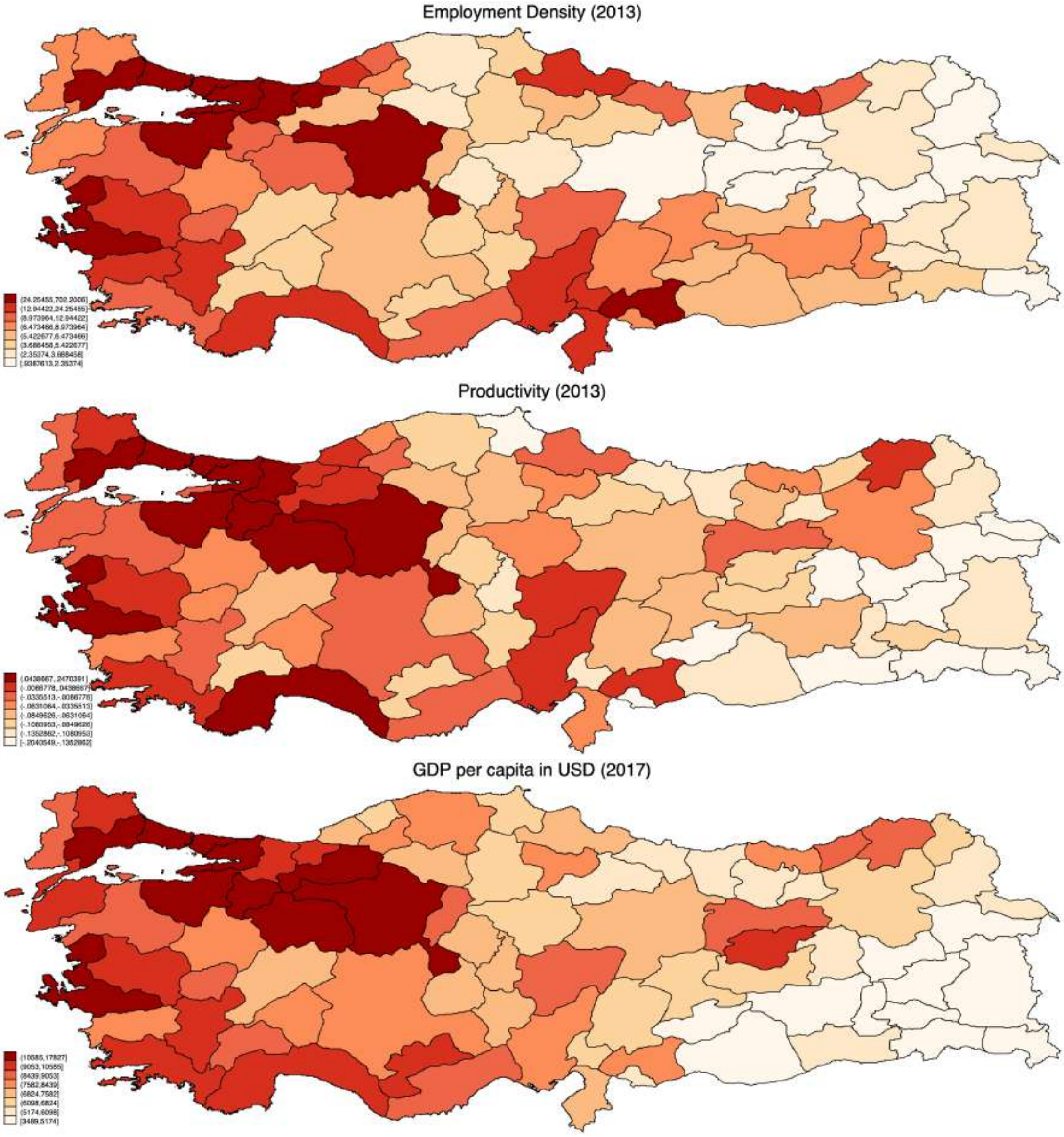
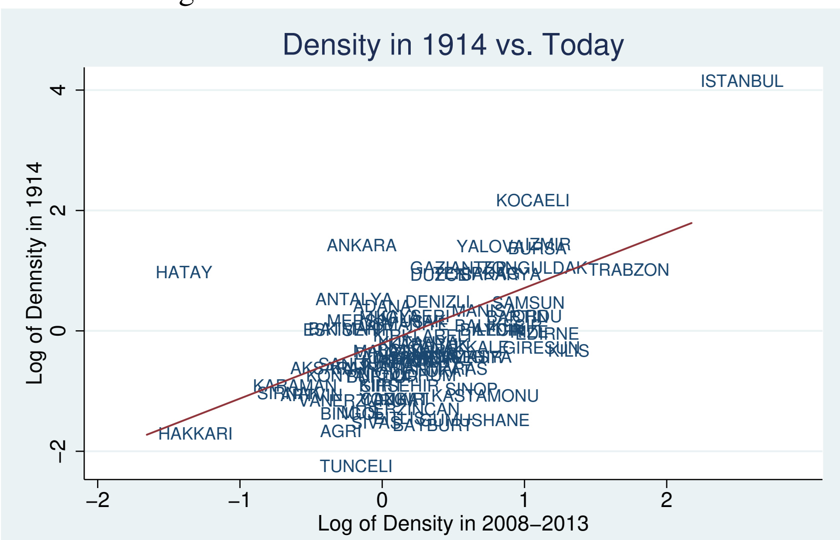


Figure 2: Instrumental Variable Relevance



Note: Each circle is proportional to the number of workers employed in the province.
Numbers are the average of 2008–2013.
Source: Karpat(1985), SGK

Table 1: OLS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Density	0.0624	0.0700	0.0506	0.0608	0.0449	0.0504	0.0614
	(0.005)***	(0.004)***	(0.005)***	(0.002)***	(0.007)***	(0.005)***	(0.007)***
Land Area		0.0433					0.0355
		(0.018)**					(0.010)***
Diversity			0.0767				0.0170
			(0.025)***				(0.020)
DMP				0.0790			0.0950
				(0.014)***			(0.023)***
FMP					0.2805		-0.0592
					(0.086)***		(0.100)
Human Capital						5.7470	3.0584
						(1.338)***	(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 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 2: First Stage Regressions, 2008-2013

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

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 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 3: 2SLS Results for 2008-2013

	OLS	Den 1914	Den 1927	Den 1935	Den 1945	Den Growth
Density	0.0624	0.0560	0.0582	0.0547	0.0537	0.0632
	(0.005)***	(0.005)***	(0.004)***	(0.004)***	(0.004)***	(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

Region effects	0.08
Individual characteristics	0.41
Firm characteristics	0.24
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 Survey (Turkstat), Karpat(1985)

Table 5: Second Stage Regressions, 2008-2013

	No Controls		Individual Controls	
	OLS	2SLS	OLS	2SLS
Density	0.0616	0.0560	0.0534	0.0512
	(0.003)***	(0.007)***	(0.003)***	(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 Survey (Turkstat), Karpat(1985)

$^{*} p < 0.10, ^{**} p < 0.05, ^{***} p < 0.01$

Table 6: 2SLS Results for 2008-2013

	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Density	0.0680	0.0641	0.0642	0.0609	0.0518	0.0541	0.0629	0.0720	0.0580	0.0635
	(0.005)***	(0.005)***	(0.002)***	(0.003)***	(0.002)***	(0.002)***	(0.005)***	(0.006)***	(0.003)***	(0.007)***
DMP			0.0904	0.1002	0.0813	0.0871	0.0816	0.1059	0.0956	0.1059
			(0.014)***	(0.020)***	(0.010)***	(0.011)***	(0.013)***	(0.021)***	(0.018)***	(0.024)***
Human Capital					5.0698	4.7589	2.7917	1.6219	2.6255	0.7702
					(0.373)***	(0.703)***	(0.715)***	(0.985)*	(0.633)***	(1.698)
Land Area							0.0286	0.0403	0.0312	0.0389
							(0.011)**	(0.012)***	(0.008)***	(0.010)***
Diversity							-0.0032	-0.0267		
							(0.014)	(0.019)		
Shores									0.0019	0.0033
									(0.002)	(0.003)
Coast									0.0399	0.0504
									(0.017)**	(0.023)**
Climate									-0.0039	-0.0061
									(0.001)***	(0.002)***
Rivers									0.0000	0.0000
									(0.000)	(0.000)
N	486	468	486	468	486	468	486	468	486	468
r2	0.87	0.86	0.92	0.91	0.95	0.95	0.96	0.95	0.97	0.96
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	0.0596	0.0581	0.0608	0.0569
	(0.004)***	(0.003)***	(0.006)***	(0.003)***	(0.006)***
Roads(KM)		0.0206			0.0116
		(0.028)			(0.021)
Village Roads(KM)		-0.0142			0.0052
		(0.017)			(0.019)
Cinema Halls			0.0085		0.0055
			(0.007)		(0.007)
Hospitals			-0.0088		0.0107
			(0.016)		(0.016)
Libraries			-0.0011		-0.0166
			(0.018)		(0.013)
January				0.0011	0.0015
				(0.002)	(0.002)
June				0.0048	0.0055
				(0.002)**	(0.003)**
Sunny Days				0.0169	0.0196
				(0.005)***	(0.007)**
Average Rain				-0.0001	-0.0000
				(0.000)	(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 excluded instruments are reported at the header of each column. 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

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$

Online Appendix for

Agglomeration Economies in a Developing Country: Evidence from Turkey

Cem Özgüzel

August 23, 2019

A Summary Statistics

Table A1: Summary Statistics for Local Variables

Variable	**Mean**	**sd**	**p10**	**p25**	**p50**	**P75**	**P90**
Density	18	66.4	1.8	3.0	5.7	11.3	24.9
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
Domestic Market Potential	3	1.6	1.3	1.7	2.4	3.3	4.9
Log Foreign Market Potential	23	0.1	23.1	23.2	23.3	23.4	23.5

Data source: Turkstat, SGK

Table A2: 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: Turkstat

B Yearly OLS results

Table B1: Yearly Regressions

	2008	2009	2010	2011	2012	2013
Density	0.0651	0.0657	0.0628	0.0614	0.0611	0.0588
	(0.004)***	(0.004)***	(0.005)***	(0.005)***	(0.004)***	(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

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

C One-step results

Table C1 reports the results from one-step regression. I regress the log of the average wage in industry s , in province p at time t, on local characteristics and specialisation. The elasticities are almost identical to the main results obtained from two-step regression. The very high F-stats and low R^2 are due to the low variability since the dependent variable varies at the industry-province-year level while the independent variables vary only at the province-year level.

Table C1: One Step Regressions, 2008-2013

	OLS	2SLS	OLS	2SLS
Density	0.0713	0.0685	0.0584	0.0667
	(0.004)***	(0.007)***	(0.003)***	(0.010)***
DMP			0.0600	0.0682
			(0.016)***	(0.017)***
HC			6.7508	4.1024
			(0.519)***	(1.932)**
Land Area			0.0399	0.0559
			(0.008)***	(0.018)***
Diversity			0.0121	0.0110
			(0.007)*	(0.018)
Specialisation			0.0759	0.0762
			(0.006)***	(0.006)***
N	175333	171755	175333	171755
r2	0.64	0.18	0.69	0.30
Cragg-Donalds Stat		513986.18		18881.31

The table reports 2SLS for the one-step of the estimation. The dependent variable is the log of average earnings in industry s, in province p at year t. The unit of observations are industry-province pairs. Excluded instruments are the same as in the second-step of the estimation in the main text. All regressions include industry and year fixed effects. Regressions are weighted by number of workers in each industry-province cell. Standard errors are clustered at region-level. Data source: SGK, Karpat(1985), Turkstat

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$

D Individual Analysis

D.1 First step individual regression

Table D1: 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)***
Vocationnal 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). The unit of observations are individuals. Regressions are weighted by survey weights. Standard errors are clustered at region-level. Data source: Household Survey (Turkstat)

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$

D.2 Robustness of the second-step

In Section 6.1.3, I exploit the individual dimension of the 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 D1 presents these results.

Table D1: Second Stage Regressions, 2008-2013

	No Controls		Individual Controls	
	OLS	2SLS	OLS	2SLS
Density	0.0580	0.0527	0.0507	0.0483
	(0.003)***	(0.007)***	(0.002)***	(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 Survey (Turkstat), SGK, Karpat(1985)

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

E 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 percent 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 percent for the nonagricultural 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 affliated to the social security. Table E1 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.[55] In Columns 3 and 4, I also include the workers who are in the informal sector.

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. In this sense, the elasticities estimated (us-ing formal employment) in the paper present the lower bound estimates of the agglomeration economimes. Although understanding the source of the difference in the elasticities is interesting, it goes beyond the scope of this paper.

[55] I 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 E1: Accounting for informal sector, 2008-2013

	OLS	2SLS	OLS	2SLS
Density (Formal)	0.0499	0.0481		
	(0.003)***	(0.005)***		
Density (Formal+Informal)			0.0638	0.0631
			(0.004)***	(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 Survey (Turkstat), SGK, Karpat(1985)

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

F First Stage of Multivariate Regression

Table F1 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 F1: First Stage Regressions, 2008-2013

	Column 2	Column 4		Column 6			Column 8			Column 10	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	LnDen	LnDen	LnDMP	LnDen	LnDMP	ISCED56	LnDen	LnDMP	ISCED56	LnDen	LnDMP
Density1914	1.2012	1.1980	0.0243	0.6511	0.0181	-0.0041	0.2469	-0.1479	-0.0008	0.3689	-0.1193
	(0.189)***	(0.189)***	(0.018)	(0.393)	(0.069)	(0.003)	(0.414)	(0.087)*	(0.002)	(0.340)	(0.080)
DenGrowth	0.0031	0.0033	-0.0004	0.0014	-0.0004	0.0000	0.0006	-0.0007	0.0000	0.0015	-0.0004
	(0.001)***	(0.001)***	(0.000)**	(0.001)	(0.000)**	(0.000)	(0.001)	(0.000)***	(0.000)*	(0.001)	(0.000)*
DMP1945		-0.4543	1.7345	0.2631	1.7325	0.0063	-0.1578	1.6712	0.0045	0.1803	1.6422
		(0.684)	(0.145)***	(0.738)	(0.143)***	(0.004)	(0.668)	(0.124)***	(0.004)	(1.138)	(0.216)***
(sum) enrolled$_{m}ale$1927				0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
				(0.000)*	(0.000)	(0.000)**	(0.000)**	(0.000)	(0.000)*	(0.000)	(0.000)
Land Area							-0.4412	-0.1727	0.0032	-0.1200	-0.0892
							(0.349)	(0.050)***	(0.002)*	(0.294)	(0.049)*
Diversity							0.9593	0.2239	0.0000	0.8798	0.2310
							(0.456)**	(0.083)***	(0.001)	(0.407)**	(0.075)***
Shores										0.0230	-0.0015
										(0.056)	(0.010)
Coast										0.9531	-0.0629
										(0.488)*	(0.115)
Climate										0.0361	0.0195
										(0.073)	(0.015)
Rivers										-0.0007	-0.0007
										(0.001)	(0.000)***
N	546	546	546	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

The table reports first-stage results. Each column reports regression results where the variable on the column head is the dependent variablem, 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)
$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$

G Additional Results: Infrastructure and Amenities

Table G1: OLS Results for 2008-2013: Infrastructure and Amenities

	(1)	(2)	(3)	(4)	(5)
Density	0.0625	0.0758	0.0564	0.0654	0.0946
	(0.004)***	(0.003)***	(0.013)***	(0.005)***	(0.014)***
Roads(KM)		0.0866			0.1142
		(0.022)***			(0.019)***
Village Roads(KM)		-0.0529			-0.0386
		(0.017)***			(0.026)
Cinema Halls			0.0363		0.0170
			(0.012)***		(0.009)*
Hospitals			-0.0422		-0.0441
			(0.026)		(0.026)*
Libraries			0.0197		-0.0173
			(0.019)		(0.016)
January				0.0011	0.0026
				(0.002)	(0.002)*
June				-0.0111	-0.0006
				(0.006)*	(0.004)
Sunny Days				0.0171	-0.0096
				(0.010)	(0.009)
Average Rain				-0.0004	0.0002
				(0.000)	(0.000)
N	486	486	486	486	486
Adj R2	0.82	0.88	0.85	0.84	0.90

The table reports OLS 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. Regressions are weighted by total employment in year. Standard errors are clustered at province-level. Data source: SGK, Turkstat, Turkish State Meteorological Service

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$