Cai, Zhengyu

Working Paper

Hours Worked of the Self-Employed and Agglomeration

GLO Discussion Paper, No. 199

Provided in Cooperation with:
Global Labor Organization (GLO)

Hours Worked of the Self-Employed and Agglomeration

Zhengyu Cai[a, b]
[a] Southwestern University of Finance and Economics
Institute of Development
Chengdu, SC 611130, China
[b] Global Labor Organization (GLO)
Email: czy@swufe.edu.cn

Abstract

This paper investigates the causal effects of agglomeration on hours worked by the self-employed. The IV estimations instrument for urbanization and localization using the minimum distance from the work Public Use Microdata Area centroid to the United States' coastlines and estimated industry share in 1930. The 2SLS results demonstrate that urbanization and localization decrease and increase hours worked of the self-employed, respectively. These results are mainly from outsourcing and competition, whereas sorting, simultaneity, and agglomeration wage effect are less likely to be influential. Additionally, only small business owners perceive the pressures of competition in localization economies. The young unincorporated self-employed are more likely to be affected by peer competitors, whereas the elder unincorporated perceive more pressures from large firms.

Keywords: Self-employed, hours worked, urbanization, localization, competition, coastlines

JEL Classification: J10, J22, J31, R11, R12

The author is grateful to Mehtabul Azam, Jon Comer, Shihe Fu, Dan Rickman, John Winters, and two anonymous reviewers for their helpful comments and suggestions. The author would also like to thank the discussants at the 52nd MVEA conference, 62nd NARSC conference, and 5th RUSE workshop for their insightful suggestions on the early versions of this paper.

1. Introduction

Rosenthal and Strange (2008) examined the relationship between agglomeration and hours worked of wage and salary labor. They find that the relationship between hours worked and localization is negative for nonprofessional workers, but positive for professional workers. This localization effect is stronger for younger workers. No significant relationship exists between hours worked and urbanization, given localization is concurrently controlled for. However, the relationship between hours worked of the self-employed and the agglomeration remains unknown. As a complement to Rosenthal and Strange (2008), this paper examines the effects of agglomeration on hours worked of the self-employed.

This paper focuses on the self-employed because a substantial amount of studies document the importance of the self-employed on regional economic growth (Stephens, Partridge, and Faggian 2013; Stephens and Partridge 2011; Acs and Armington 2006; Glaeser, Rosenthal, and Strange 2010; Rocha 2004). An influential feature of small businesses as regional economic engines is job creation (Henderson and Weiler 2010; Baptista, Escária, and Madruga 2008; Glaeser, Kerr, and Ponzetto 2010; Parker 2004). Longer hours worked by the self-employed may secure their survival to increase the probability of success (Portes and Jensen 1992; Portes and Zhou 1996; Douglas and Shepherd 2000), creating additional jobs. Although the evidence is mixed regarding whether the self-employed are more likely to locate in clusters (Parker 2004), Delgado, Porter, and Stern (2010) find that clusters increase entrepreneurship activities and contribute to start-up firm survival. Furthermore, J.

Henderson and Weiler (2010) show that the impact of entrepreneurship on job growth is greater in the areas that are more urbanized. Given that the literature on labor supply decision of the self-employed is insufficient (Parker 2004), studying whether and how agglomeration would affect hours worked of the self-employed is an arresting endeavor.

The self-employed tend to be relatively versatile in their skill sets and multitask in their daily work. Thus, hours worked by the self-employed are not necessarily comparable with those by salaried workers. Therefore, this paper only focuses on the self-employed and answers the aforementioned question that whether and how agglomeration affects hours worked by using the Integrated Public Use Microdata Series (IPUMS. Ruggles et al., 2010).

After controlling for a large set of amenity measures constructed by using geographic information system (GIS) data at the Public Use Microdata Area of work (work PUMA) level and other covariates, the ordinary least squares (OLS) results show that urbanization is negatively correlated with hours worked of the self-employed, and the relationship for localization is positive. The instrumental variable (IV) estimations reduce the concerns of the endogeneity issue and confirm the OLS results.

Compared with Rosenthal and Strange (2008), like their paid counterparts, the self-employed work longer hours in more localized areas. They explain this positive relationship using the urban rat-race effect, which is essentially derived from competition at an occupational level. Similarly, this paper finds that the positive relationship for the self-employed also comes from competition, but at an industrial level. This paper also shows that only small business owners perceive the pressures of competition in localized economies. The young unincorporated self-employed are more likely to be affected by peer competitors, and the elder unincorporated perceive more pressures from large firms.

The other results of the self-employed are somewhat different from Rosenthal and Strange (2008), which is less surprising because the analytical samples are from two distinct groups of workers. First, in most subsamples, urbanization significantly decreases hours worked of the self-employed. Second, the effects of urbanization and localization are stronger for elder self-employed workers.

The remainder of this paper is organized as follows: Section 2 provides a conceptual framework. Section 3 describes the empirical framework used in this paper and discusses identification issues. Section 4 introduces the data and variables. Section 5 shows the empirical results. The last section concludes.

2. Conceptual Framework

A critical advantage of being self-employed is a flexible work schedule (Boden 1999; Loscocco 1997). Yet, Portes and Jensen (1992) argue that extra work effort is inherent to self-employment. This phenomenon has been confirmed by censes and survey results: the self-employed work significantly longer hours than paid workers (Portes and Zhou 1995, 1996; Bailey and Waldinger 1991; Baines and Gelder 2003; Parker 2004; Carrington, Mccue, and Pierce 1996). The self-employed are willing to work additional hours because they have less restrictions regarding their choice of work hours than their paid counterparts, when well compensated (Portes and Zhou 1996).

Notably, the weekly hours worked gap between self-employed and paid workers is shrinking over time (Aronson 1991; Rees and Shah 1994; Moralee 1998; Parker 2004). Parker (2004) believes that this trend is possibly due to the increase in female self-employed, who are more likely to work part-time. Given the literature on the variation in labor supply decisions across employment types, this paper focuses on the variation in hours worked within an employment type, namely, self-employment. To partially explain such variation, this paper considers agglomeration.

Several possible factors could lead to the hours worked of the self-employed to vary in agglomeration. The first pair of channels are competition and outsourcing. The benefits of localization attract additional businesses into a cluster and raise competition within an industry. To smooth the benefits, the self-employed tend to increase their work intensity when competition increases.[1] By contrast, specialization in urbanized areas leads to less competition across industry boundaries, which reduces hours worked. Furthermore, Helsley and Strange (2011) document that the less thick input markets in localized areas increase the completion time of entrepreneurial

[1] From a perspective of spatial equilibrium, the self-employed can relocate or change industries to prevent being worse-off because of competition. However, Greenwood et al. (1991) and Clark et al. (2003) argue that spatial equilibrium is a long- rather than short-run phenomenon. Thus, a worker will move to a second-best location after competition swallows all the additional benefits (referring to the second-best location) brought by localization, which takes time. Additionally, spatial equilibrium is only valid for marginal movers (Gyourko, Kahn, and Tracy 1999; M. E. Kahn 2006; Krupka and Donaldson 2013; Roback 1988). Some workers do not move because of friction (Cai 2018). Thus, an increase in hours worked is one means to overcome competition, even when considering spatial equilibrium.

projects.[2] Conversely, the thicker and more diverse factor markets in urbanized areas reduce the completion time, because workload can be more easily outsourced in thicker markets. Therefore, urbanization is more likely to decrease work intensity of the self-employed, and localization could influence them to increase their work hours.

Sorting and simultaneity could be the second pair of channels. Workers with a higher work intensity might sort into certain areas or industries based on unobservable characteristics of the areas, industries, or individuals (Portes and Zhou 1996; Parker 2004). These unobservable characteristics could influence the work intensity pattern of the self-employed in agglomeration. For instance, people who prefer longer hours worked may sort into self-employment to earn full compensation for their efforts. Additionally, if agglomeration increases productivity, the self-employed, who tend to work longer hours, would sort into a denser area for increased compensation. Hardworking self-employed individuals could also be attracted by urban amenities. Moreover, reverse causality may also exist. Longer hours worked could increase competition, increase human capital level, and thicken markets, which generates agglomeration. If such simultaneity exists, the descriptive relationship between hours worked and agglomeration could be biased.

Another channel could be an agglomeration wage effect associated with urbanization economies.[3] Many studies have substantiated that urbanization increases

---

[2] Although inter-area trade can ameliorate less thick markets, frictions, such as transaction costs, could increase project completion time.

[3] Most studies have documented the agglomeration wage effect in an urbanization context rather than localization (Ciccone 2002; Rosenthal and Strange 2004; Brülhart and Sbergami 2009; Combes et al. 2010). The online Appendix provides the results of an agglomeration wage effect test using the sample

productivity and wages (Ciccone 2002; Rosenthal and Strange 2004; Brülhart and Sbergami 2009; Combes et al. 2010).Standard labor supply models decompose hours worked responses to such an urban wage premium into two parts: the substitution effect (SE) and income effect (IE). If the SE of a higher wage dominates the IE, the higher the wage the more labor will be supplied. If the IE dominates the SE, a negative relationship exists between the wage and hours worked (Parker 2004). Controlling for work metropolitan statistical area (MSA) fixed effects would include the agglomeration wage effect on work-hour variation, because many studies have found the urban wage premium at the scope of the MSA (Glaeser and Maré 2001; Yankow 2006).

3. Empirical Framework

3.1. OLS Estimation

The relationship between agglomeration and hours worked is estimated by an OLS first. The empirical model is as follows:

$$\begin{aligned}\log(y_{icd}) = \alpha \log(Urbanization_c) + \beta \log(Localization_{cd}) \\ + \boldsymbol{X}_{icd}\boldsymbol{\gamma} + \boldsymbol{A}_c\boldsymbol{\theta} + \tau_d + \pi_m + \varepsilon_{icd}\end{aligned} \quad (1)$$

where *i* indexes individual observations, *c* denotes work PUMA, *d* denotes industries, and *m* denotes work MSAs.[4] The dependent variable is the log of hours worked. $Urbanization_c$ is the population density of a work PUMA. $Localization_{cd}$ is

---

of this paper. The results confirm that the agglomeration wage effect comes from urbanization rather than localization. Thus, the agglomeration wage effect could be a channel through which urbanization affects hours worked, but not for localization.

[4] All the empirical steps, including all the regressions, use personal sampling weights to ensure the results are nationally representative.

measured by the industry-specific self-employment density of a work PUMA.[5] $\boldsymbol{X}_{icd}$ is a standard set of demographic characteristics. $\boldsymbol{A}_c$ is a set of amenities, which is included because amenity shocks may influence hours worked. Controlling for regression-adjusted wages and rents helps ameliorate the shocks (Winters 2013); however, these controls are potentially endogenous (Winters 2013; S. Kahn and Lang 1991; Portes and Zhou 1996). Therefore, amenities are directly included in the empirical specifications. τ_d is an industry fixed effect. π_m is work a MSA fixed effect. ε_{icd} is an error term.

3.2. Identification Issues

As discussed in Section 2, sorting and simultaneity could still bias the estimates of the coefficients in Model (1), even though the model controls for a large set of observable variables. To reduce the concerns of sorting and simultaneity, IV estimation is employed. The minimum distance between the work PUMA centroid and the United States' (US) shoreline[6] and industry share in 1930 are used as the instrumental variables for urbanization and localization in Model (1).

For a preview of the instruments, a preliminary discussion on the relevance and exogeneity conditions is provided. The relevance condition requires a valid

---

[5] The self-employed are not like most employees, who can be identified with a certain occupation; they could have several occupations simultaneously. Industry is a better scope to identify the self-employed. Thus, the self-employment density of a given industry is used in this paper, rather than a given occupation, to measure localization. See Appendix A3 for the details on constructing the independent variables of interest.

[6] It includes the shorelines of Atlantic Ocean, Pacific Ocean, Gulf of Mexico, and Great Lakes.

instrumental variable to be strongly correlated with the instrumented variable. For the first instrument, the strong correlation between population density and distance to shorelines is documented in the economic, geographic, and anthropologic literature. Specifically, Beeson, DeJong, and Troesken (2001) use the 1840–1990 US county-level census data and find a positive correlation between population density and ocean proximity. In their renowned paper, Rappaport and Sachs (2003) find a similar relationship between population concentration and proximity to ocean and Great Lakes coasts. In the coastal hazard literature, alike correlations between population density and proximity to shorelines have been recorded (Small, Gornitz, and Cohen 2000; Small and Nicholls 2003). In the anthropologic literature, likewise, population is found to be heavily distributed in costal zones and diminishes with distance to coastlines (Small and Cohen 2004). In summary, coastal areas are much denser; this phenomenon is also true in the analytical sample according to Figure 1, which shows the population density for each work PUMA in 2000.

The first instrument is shown in Figure 2. The minimum distance from work PUMA centroid to shoreline (in red) is calculated, that is, the lengths of those orange lines. The different colors of work PUMAs indicate the distance differentials: a darker color indicates a further distance to the shoreline, and a lighter color indicates the opposite. Comparing Figure 2 with Figure 1, a similar pattern is observed.

The second instrument, industry share in 1930, is arguably reliable because the historical (long lagged) variables are often used in the literature as instrumental variables (Ciccone and Hall 1996; Combes et al. 2010). The historical variables are relatively exogenous to current economic outcomes. To impute industry shares in 1930 at the work PUMA level, the employment by industry in 1930 is calculated at the county level.[7] Next, the county level data is converted to work PUMA level by using the allocation factor from MABLE/Geocorr2K: Geographic Correspondence Engine with Census 2000 Geography available from the Missouri Census Data Center, which is used as the estimated allocation factor in 1930. Lastly, industry shares are calculated for each work PUMA.

For a relatively formal test of the relevance condition, the correlation coefficients between log urbanization and log minimum distance to shoreline, and between log localization and log 1930 industry share are calculated. Figures 3 and 4 show the raw correlation coefficients are -0.6038 and 0.2525, respectively, which are substantial. Additionally, the formal first stage weak identification tests in Section 5.2 show that the instruments are less likely suffering from the weak instrument problem.

The exogeneity condition requires that a valid instrument is not causally related to the dependent variable in the second stage. In other words, looking at the first instrument, the minimum distance to shoreline should only affect hours worked through agglomeration. One obvious concern, however, is that the minimum distance to shoreline affects productivity and, hence, hours worked because it also measures the accessibility to ports and harbors, which could increase productivity.

One possible solution is directly controlling for the work PUMAs with ports

---

[7] The data is from the IPUMS 1930 5% population sample. IND1950 is available in this sample, which identify industries by a 1950s basis. Because IND1950 is also available in the IPUMS 2000 sample, the consistency is guaranteed.

and harbors. In the preferred specification, dummies for coastal work PUMAs are controlled for, which includes the work PUMAs with ports and harbors. Another strategy is excluding all port work PUMAs from the sample. Yet, due to the untestable nature of the exclusion restriction, ruling out all potential sources of endogeneity, like measurement errors, is difficult. A relatively feasible method to mitigate this concern is to include additional instruments and conduct an overidentification test to observe if the instruments are all exogenous. However, the efficacy of the overidentification test is based on the maintained hypothesis that the model is exactly identified, which is untestable in the first place. Thus, the concern of violating the exclusion restriction can be mitigated but not eliminated.

4. Data and Variables

This paper uses a 5% national random sample of the population in 2000 that covers the contiguous 48 US states. Only male, full-time, self-employed workers aged 30 to 59 who work 35 hours or more per week are included.[8] To illustrate the heterogeneity by age and education, the sample is subdivided into three groups: young, middle-aged, and elder worker groups aged from 30 to 39, 40 to 49, and 50 to 59, respectively. For each subsample, these groups are further divided into two educational groups: a high school degree or less, and a college degree or more. College dropouts are excluded from the sample to ensure the division is sharp.[9]

---

[8] People aged 30–59 comprise approximately 80% of the full-time, self-employed workforce in the sample. See Appendix A1 for the details on how to identify the self-employed in the data.

[9] College dropouts are a special group, compared with the other two groups. They cannot be integrated

The empirical models use usual hours worked per week in the previous year as the dependent variable.[10] Kahn and Lang (1991) find that using actual hours worked rather than desired hours to measure working hours of the self-employed is satisfactory. The deviation between actual and desired hours is not much for the self-employed, because they have less restrictions regarding choosing their hours worked and are well-compensated for working longer hours (Portes and Zhou 1996).

All estimated models in this paper control for a standard set of demographic attributes, including educational attainment, a dummy of presence of children, dummies of marital status, a quartic polynomial of age, dummies of race, years of residency in the US, and travel time to work.

Amenities are extracted from different sources and constructed at the work PUMA level, including violent crime, property crime, precipitation, January temperature, July temperature, elevation, minimum distance to the nearest river or lake, heating degree days, cooling degree days, dew points, direct solar irradiance, and dummies for the coastal work PUMAs of the Atlantic Ocean, Pacific Ocean, Gulf of Mexico, and Great Lakes.[11]

Table 1 shows the summary statistics for the analytical sample. A typical male, self-employed worker is approximately 45 years old and works approximately 51 hours per week. He is highly likely to be a married, non-Hispanic, white man with

---

into any other groups due to different behavior patterns. The empirical results show that most estimates for college dropouts are trivial (these results are provided upon request).

[10] Appendix A2 provides the detailed information of the dependent variable.

[11] Detailed information about data source and variable construction of amenities is provided in Appendix A4.

children present at home but less likely to have a college degree. He is working in an area of 3,134 people per square mile, and 9 people per square mile who are running businesses in the same industry. The top three industries he possibly works in are construction, agriculture, and medical plus other health services (except hospitals).[12]

5. Empirical Results

5.1. OLS Results

This section reports the OLS estimates of Model (1). Firstly, full sample results are presented in Table 2. Column (1) shows that the self-employed work less hours in areas that are more urbanized. On average, a 1% increase in urbanization is significantly associated with 0.0176% less hours worked by the self-employed. This negative relationship is consistent with the discussion in Section 2: it is easier for business owners to outsource parts of their projects from a more diverse and thicker market in a highly urbanized area.

Since work MSA fixed effects are controlled for in all the regressions, such a negative relationship between hours worked and urbanization is less likely to have come from the agglomeration wage effect; this is consistent with Parker, Belghitar, and Barmby (2005), who indicate that the labor supply of the self-employed is not associated with wage per se. Additionally, this negative effect is less likely to be caused by a longer commute time and better urban amenities because the estimates are very similar with and without controlling for commute time and work MSA fixed

[12] The shares are 24.15%, 10.43%, and 6.74%, respectively.

effects.[13] To improve the illustration of this negative correlation between urbanization and hours worked of the self-employed, Manhattan, New York, NY, is taken as an example, which is the densest work PUMA in the analytical sample.[14] The urbanization measure of Manhattan is approximately 66,942 people per square mile, whereas the sample average is approximately 3,134 people per square mile (Table 1). The average hours worked of the self-employed in Manhattan is 50.541 hours per week, which is below the sample average of 50.877 hours per week.

Column (1) also indicates that the self-employed work more hours in more localized areas. Specifically, a 1% increase in localization is significantly correlated with 0.0120% more average hours worked of the self-employed. This positive relationship may be caused by a rivalry within industries. A place is more localized with an industry, meaning more people are running businesses in the same industry. Thus, people would work more hours to survive in such a rivalry.[15] Taking Manhattan as an example again, it is the most localized work PUMA nationwide for the banking and credit agencies industry.[16] There are approximately 37 people per square mile running their own business in this industry in Manhattan, and they work on average 50.363 hours per week. The work PUMA with average localization of the banking and credit agencies industry is Orange County (excluding Irvine),

[13] Results are provided upon request.

[14] Work PUMA ID for Manhattan, New York, is 3603800.

[15] The positive relationship could also come from a demand effect. A Bartik shift-share is included in a robustness check. See Section 5.3 for details.

[16] Industry code of banking and credit agencies is 716.

California.[17] Approximately two people per square mile own a business in this industry, and their average hours worked is 49.345 hours per week. There is greater than a one-hour difference per week between the hours worked by the self-employed of the banking and credit agencies industry in Manhattan and Orange County (excluding Irvine). As discussed, this difference may be caused by perceived severer competition.

To control for potential self-employment competition across industry boundaries, Column (2) of Table 2 shows the results with an additional variable of the self-employed share at the work PUMA level; however, the estimate of this variable is insignificant. The coefficients of interest are not qualitatively changed.

Self-employed and paid workers are two distinct groups of workers, but industry forces may affect them similarly. Column (3) of Table 2 reports the results with an additional variable of the self-employed share within industries at the work PUMA level, which ameliorates the possible effect. The estimated coefficient of this variable is significant at the 10% level; however, it does not change the robustness of the coefficients of interest. When these two self-employed share variables are simultaneously included in the regression, none of the corresponding estimates is significant, as shown in Column (4). Thus, the specification in Column (1) is good for the intended purpose and will be used as the main specification hereafter.

Table 3 reports the OLS estimates of Model (1) by education and age. All estimates for urbanization are negative. The estimates are significant and substantial,

[17] Work PUMA ID is 606890.

especially for the lower educated group. The estimated elasticities for the lower educated, self-employed varies from -0.0281 to -0.0173. For the higher educated group, the range is from -0.0191 to -0.0075. The estimates for localization are positive for all educational and age groups. The estimated effects are still larger in magnitude for the lower educated group, which is reminiscent of some agglomeration literature that has documented that less educated workers benefit more from externalities (see Winters (2013) as an example).

In terms of age, the patterns are different across educational groups. For the lower educated group, the absolute effects of urbanization and localization are greater for younger workers. For the higher educated group, both effects are greatest in absolute magnitude for elder workers and smallest for middle-aged workers; yet, the effects are not causal. Next, the IV estimation is used to reduce the concern of the endogeneity issue as discussed in the previous sections.

5.2. IV Estimation Results

Table 4 shows the two-stage least squares (2SLS) results with the minimum distance from the work PUMA centroid to the coastline and industry share in 1930 as the instruments. All the estimated coefficients of interest increase in their magnitudes, but no qualitative change is observed. Urbanization reduces the hours worked of the self-employed, and localization leads to the opposite.[18]

[18] Under certain assumptions, an IV estimate provides a local average treatment effect (LATE) (Angrist and Imbens 1995; Angrist, Imbens, and Rubin 1996; Heckman, Urzua, and Vytlacil 2006; Imbens and Angrist 1994); however, it is less likely to interpret the IV estimates as LATE in this paper. LATE is the

The effects are still larger for the lower educated self-employed. A 1% increase in urbanization reduces the hours worked of the lower and higher educated self-employed groups from 0.0477% to 0.0694% and from 0.0173% to 0.0451%, respectively. A 1% increase in localization increases the hours worked of the lower and higher educated self-employed groups by 0.0388% to 0.0585% and 0.0199% to 0.0271%, respectively.

All the effects increase with age. Regardless of education level, higher urbanization makes the elder self-employed work less, but higher localization forces them to work more than their younger counterparts.

According to the endogeneity tests, the null hypothesis that urbanization and localization are exogeneous in the OLS regressions can be rejected for all subsamples at the 5% level. After mitigating the endogeneity issue by using IV estimations, the relationship between urbanization and hours worked remains negative; and it remains

---

average treatment effect only for compliers. In this paper, the definition of complier is somewhat different from a traditional context. For simplicity, we only focus on urbanization (population density). In this case, minimum distance to a shoreline is treatment, and population density is outcome. LATE is usually discussed in a discrete treatment context, but it can be generalized to a continuous treatment setting as in this paper. Assume the closer to a shoreline the more treated. Thus, the "compliers" refer to the areas close to a shoreline with a large population density, or the areas far from a shoreline with a small population density. The IV estimates only provide the effect on the people in these areas. However, the treatment here is not directly assigned to a person but an area. Formation of a city or population density that we observed is based on the collective behaviors of a group of people rather than an individual. Thus, it is difficult to define a real complier (or defier, always-taker, never-taker) at individual level, because individual force cannot change the role of an area. Assuming we have a complier referring to a person (of course we cannot define her in this context), this person lives in an area close to a shoreline but with a small population density. By definition, this area is a defier or never-taker. Yet, the role of the person is not changed by the role of the place she lives in. In other words, people choose their locations without considering whether the places are treated or not. Therefore, it would mitigate the concern that IV estimates are LATE in this paper.

positive for localization and hours worked. Both effects become even greater in absolute magnitudes. Therefore, the IV estimation results imply that the relationship between agglomeration and hours worked by the self-employed is less likely to have resulted from sorting and simultaneity, but mainly comes from competition and outsourcing. The boost of the OLS estimates may be also caused by potential measurement errors in the two agglomeration variables.

The significant and considerable first stage estimates indicate that, as expected, minimum distance to the shoreline is negatively correlated to agglomeration, and historical industry share is positively associated with agglomeration. In this case, given that the Stock-Yogo weak instrument test critical value based on a 5% size distortion of a 5% Wald test is 7.03 (Stock and Yogo 2005), all the first stage Kleibergen-Paap Wald F statistics (Kleibergen and Paap 2006) are sufficiently large to provide solid evidence that the instruments are less likely to suffer from the weak instrument issue.

5.3. Robustness Checks

Several robustness checks are conducted to reinforce the credibility of the main results. Firstly, as discussed in Section 3.2, the first instrument is less likely to violate the exclusion restriction if it can be shown that accessibility to ports and harbors is not correlated with hours worked. One possible solution is directly controlling for the work PUMAs with ports and harbors. Aggressively, in this paper's preferred specification, dummies for coastal work PUMAs are controlled for, including all the work PUMAs with ports and harbors.[19] The instruments are not weak, according to the weak identification statistics with the coastal dummies in these regressions (Table 4). Another strategy is excluding all the port work PUMAs from the sample.[20] Panel A of Table 5 reports the IV estimation results. The estimates are not qualitatively different from those in the full sample specifications, ruling out the alternative channel.

Although it is not possible to eliminate all sources of endogeneity, a relatively feasible and useful attempt is to include additional instruments and conduct an overidentification test to observe whether the instrumental variables are all exogenous. Specifically, the population density in 1930 is employed as an additional instrument. Panel B of Table 5 shows that the estimates are robust. More importantly, all the Hansen J overidentification tests fail to reject the null hypothesis, that is, all the instruments are exogenous. Yet, as aforementioned, these results should be interpreted cautiously because the overidentification test is based on the maintained hypothesis that the model is exactly identified.

To warrant that the localization effect is mostly from competition and not a demand effect, a standard Bartik shift-share is constructed and included in the regression models. A Bartik shift-share is used to generate exogenous labor demand shocks (Bartik 1991; Blanchard et al. 1992); specifically, to construct the Bartik shift-

[19] Inland ports are not considered.

[20] Excluding inland ports, there are 84 work PUMAs with at least 1 ports. The data comes from U.S. Geological Survey, 201406, USGS Small-scale Dataset - Ports of the United States 201406 Shapefile: U.S. Geological Survey.

share, 1990 is chosen as the base year for data availability. The same three-digit industry code used to construct the localization measure is used to construct the Bartik shift-share. Panel C of Table 5 reports the results. All the estimates of the Bartik shift-share are insignificant except the youngest group of the lower educated; however, the estimates of interest remain robust to the main results. Thus, the positive effect of localization is less likely to come from a demand effect.

The sophisticated correlation between urbanization and localization shows that the partial R squared is 0.4688. Thus, the partial correlation is approximately 0.7, implying the possibility of collinearity. To reduce the concern, a nonlinear transformation on the localization measure is performed:

$$\log(y_{icd}) = \alpha \log(Urbanization_c) + \beta \frac{\log(Localization_{cd})}{\log(Urbanization_c)} + \boldsymbol{X}_{icd}\boldsymbol{\gamma} + \boldsymbol{A}_c\boldsymbol{\theta} + \tau_d + \pi_m + \varepsilon_{icd} \quad (2)$$

where the quotient $\frac{\log(Localization_{cd})}{\log(Urbanization_c)}$ measures relative localization. Panel D of Table 5 shows no qualitative change in the results. Another strategy to mitigate the collinearity concern is to replace the urbanization variable (and the other variables measured at the work PUMA level) by the work PUMA fixed effects. Panel E of Table 5 shows that the estimated localization effects are still highly robust.

Panel F of Table 5 reports the results with standard errors clustered by the work MSA, which is a higher geographic level than the work PUMA. However, compared with Table 4, little change in the standard errors is observed, and the significance levels of the results do not change at all. Therefore, clustering by the work PUMA is sufficient for the purpose.

The last two panels of Table 5 present the results for immigrants and natives separately. For the immigrant sample, the general pattern of the results does not change much, except that both effects are larger for the younger group. Certain estimates are noisy, possibly due to the small number of observations. For the native sample, the results are very similar with the main results, except for the lower precision.

5.4. Heterogeneity: Peer Effect and Counterpart Effect

Most self-employed workers run small businesses. Due to the heterogeneity of industrial organization, competitive industries consist of many small firms competing against each other. If an industry is dominated by a few giants, then a small business faces competition from peers and large firms, whereas a large firm competes with its peers in addition to being challenged by innovative startups. Examining how the different sources of competition influence the different types of self-employed businesses in the analytical sample is arresting.

The incorporated self-employed and their unincorporated counterparts are essentially different (Levine and Rubinstein 2017). Incorporated firms outperform small businesses in many aspects. In this exercise, the incorporated are used as the proxy for large firms, and the unincorporated for small businesses.[21]

To observe how peer and counterpart effects are associated with hours worked of the self-employed, the localization variable is replaced by two new variables. The

[21] This is an approximation due to lacking firm level data. The caveat is that some workers may work in the same firm. This probability is assumed to be somewhat low in the sample.

first variable measures peer effect:

$$Peer\ Effect_{cd|w} = \frac{Self\ Employment_{cd|w}}{\sum_c Self\ Employment_{cd|w}}$$

where w denotes incorporated or unincorporated. For the incorporated, this variable is constructed as the industry-specific national share of incorporated self-employment in a work PUMA and measures the concentration of large firms. Similarly, for the unincorporated, the peer effect variable is the industry-specific national share of unincorporated self-employment in a work PUMA, measuring the concentration of small businesses.

The second variable measures the counterpart effect. Since the peer effect variable for the incorporated measures the concentration of large firms, it can be used as the counterpart effect measure for the unincorporated sample:

$$Large\ firm\ Effect_{cd} = \frac{Incorporated\ Self\ Employment_{cd}}{\sum_c Incorporated\ Self\ Employment_{cd}}$$

Similarly, the counterpart effect variable for the incorporated sample is the peer effect measure for the unincorporated:

$$Small\ firm\ Effect_{cd} = \frac{Unincorporated\ Self\ Employment_{cd}}{\sum_c Unincorporated\ Self\ Employment_{cd}}$$

Table 6 reports the OLS results.[22] For the unincorporated, peer competition causes young people to work more hours regardless of education. For elder, unincorporated self-employed, the estimates are positive but mostly not significant, which may be because experience offsets hours worked under peer pressure from

[22] Experimentation with 2SLS is also performed. Minimum distance from the work PUMA centroid to the shoreline is used to instrument for urbanization; however, the results are similar with the OLS results. The endogeneity tests also suggest that urbanization is less likely to be endogenous in this specification.

small businesses. Large-firm pressure on the unincorporated is more significant for the elder workers. This phenomenon may be because the opportunity cost of switching status is higher for elder small business proprietors under the large-firm pressure; thus, they must work more hours to maintain their current businesses. For the incorporated, neither peer pressure nor counterpart effect significantly affects hours worked, although most estimates are positive.

In summary, the positive spillover effect of localization on hours worked of the self-employed is mostly from the pressure of competition, and the pressure of competition is mainly perceived by small business owners. Young proprietors are more likely to be affected by peer competitors, whereas elder small business owners perceive more pressures from large firms.

6. Conclusion

This paper documents the effects of agglomeration on the labor supply of the self-employed as a complementary study to Rosenthal and Strange (2008). The OLS results show that the self-employed work less hours in urbanized areas while working more hours in localized areas. The negative relationship may come from outsourcing in thick and diverse markets. The positive relationship may result from competition within industry.

This paper uses minimum distance from the work PUMA centroid to the US coastlines and industry share in 1930 as instrumental variables for agglomeration to reduce the endogeneity concern. The 2SLS results qualitatively confirm the OLS results. Since the 2SLS estimates are larger in absolute magnitude than the OLS estimates, sorting and simultaneity are less likely to be influential; however, measurement error cannot be eliminated.

The positive spillover effect of localization on hours worked of the self-employed mostly comes from the pressure of competition. Evidence is observed that the pressure of competition is mainly perceived by small business owners. Young proprietors work more hours in localization economies influenced by peer competitors, whereas elder small business owners work more because of perceived pressures from large firms.

In summary, the empirical findings in this paper suggest that the variation in hours worked of the self-employed partially depends on the types of agglomeration. Diversity caused by urbanization externalities provides additional chances to decrease extra working time. Competition in localization economies influences the self-employed to work additional hours, especially small business owners.

This paper contributes to the literature by establishing a causal relationship between agglomeration and hours worked of the self-employed; however, caveats remain. Due to the dearth of the firm level data, the identification of the peer effect and counterpart effect of competition may be questionable; thus, future work can address this deficiency.

Appendices

A. Sources and Construction of Variables

A1. Identifying the Self-Employed

The class of worker variable (classwkr in IPUMS) is used to identify the self-employed. The detailed variable of the class of worker (classwkrd in IPUMS) is used to identify large firms and small business owners as incorporated and unincorporated. The people either working in non-MSAs and/or living in non-MSAs are included in the sample by recoding the identifier with the work/residential state variables (pwstate2/statefip in IPUMS, respectively). The sample excluding people working in non-MSAs is also tested, the results are omitted due to similarity.

A2. Dependent Variable

The empirical models use usual hours worked per week in the previous year (uhrswork in IPUMS) as the dependent variable, and experimentation with annual hours worked was also conducted. To obtain the annual data, hours worked per week was used, multiplied by weeks worked (wkswork1 in IPUMS). The results are similar; however, concerns exist regarding the annual data. The description from IPUMS is as follows, "For employers, WKSWORK1 covers all weeks that the business or farm was in operation, even if the employer was absent." Thus, this data does not represent the actual hours, which may possibly result in large measurement errors.

A3. Independent Variables of Interest

To calculate the population density for a work PUMA, population and land area data of the work PUMA is required. Population and land area data for the 2000 work

PUMA are extracted from the Missouri Census Data Center (http://mcdc.missouri.edu/websas/geocorr2k.html). Unfortunately, the data are only available for residential PUMAs. Since the work PUMA is coded differently from the residential PUMA, population and land area data for the work PUMA is calculated from the corresponding residential PUMA by matching their codes. The relationship between residential PUMA and work PUMA is provided by the IPUMS table (https://usa.ipums.org/usa/volii/00pwpuma.shtml#5percent).

The localization measure is constructed within the sample. Self-employment in each industry is calculated for each work PUMA and adjusted by personal weight. Next, using this industry-specific self-employment, we divided by the geographic area of the work PUMA to obtain the localization measure. The variable of IND1950 in the IPUMS is used to identify industries and is a three-digit identifier: 124 three-digit industry categories are used to construct the localization measure. Next, IND1950 is recoded to a one-digit identifier based on broader groups, and 11 one-digit industry categories are included as industry fixed effects.

A4. Amenity Variables

The crime data is from the Uniform Crime Reporting Program Data ("United States Department of Justice. Federal Bureau of Investigation. Uniform Crime Reporting Program Data [United States]: County-Level Detailed Arrest and Offense Data." 2000), and this paper uses its county-level detailed arrest and offense data from 2000, which covers all the counties in the US states, except those in Wisconsin, Illinois, Florida, and the District of Columbia (DC). Violent crime and property crime data for

Wisconsin, Florida, and DC come from the US Counties website, but the Illinois data is still missing. Crime data for Illinois' counties are extracted from the Illinois County website. This county-level data is converted to the PUMA level by using the allocation factor from MABLE/Geocorr2K: Geographic Correspondence Engine with Census 2000 Geography available at Missouri Census Data Center. Next, the geocodes of the PUMA are recoded for the work PUMA.

Precipitation and dew points data are obtained from the old version of PRISM (PRISM Climate Group, Oregon State University, http://oldprism.nacse.org, extracted on September 19, 2015.). They are 30-arc-second (800 meters) gridded raster data. For precipitation, 30-year (1971–2000) annual average data is used. For dew point, 10-year (1991–2000) annual average data is constructed. Boundary file for work PUMAs is available at the IPUMS website (https://usa.ipums.org/usa/volii/00pwpuma.shtml). Next, mean precipitation and dew points data for each work PUMA are calculated by GIS software.

January and July temperatures are extracted from the current version of PRISM (PRISM Climate Group, Oregon State University, http://prism.oregonstate.edu, extracted on September 20, 2015.). Monthly raster data at a 4-kilometer grid cell resolution from 1981–2000 is used to construct the 20-year average data. Next, the mean January and July temperatures for each work PUMA are calculated using the work PUMA shapefile.

Elevation data are extracted from the hole-filled seamless SRTM data V4.1 distributed by the International Centre for Tropical Agriculture (CIAT) (Jarvis et al.

2008). The data source is the Shuttle Radar Topography Mission (SRTM) of the National Aeronautics and Space Administration (NASA) and available from the U.S. Geological Survey (USGS). The original SRTM data are available at 1 and 3 arc-second grid cell resolutions but with small voids. The data distributed by the CIAT filled the voids using interpolation methods with 3 arc-second grid cell resolution (approximately 90 meters). The Global 30 Arc-Second Elevation (GTOPO30) data are also experimented with, and the results are similar. Considering its lower resolution, results from the GTOPO30 are not presented in this paper. The average elevation for each work PUMA is calculated using the shapefile.

River centerlines and lake shapefiles with 1:10 million scales are available at the Natural Earth website (http://www.naturalearthdata.com/downloads/10m-physical-vectors/). The global datasets are merged with the North America supplement datasets. River centerlines and lakes in the US are clipped from the global merged data by the work PUMA shapefile. The Great Lakes are excluded from the dataset, because they are used to construct the coastline. The minimum distance from the work PUMA centroid to the nearest river or lake is calculated by GIS software.

Solar irradiance, heating degree days, and cooling degree days are retrieved from the National Renewable Energy Laboratory (NREL) under the U.S. Department of Energy. Direct normal irradiance (DNI) at a 10-kilometer resolution for the contiguous 48 states is used to construct the average data for each work PUMA. Heating and cooling degree days are derived by the Solar and Wind Energy Resource Assessment (SWERA) from NASA's Surface meteorology and Solar Energy (SSE) dataset. One-degree cell resolution GIS data are available at the NREL website. Next, a similar approach is applied to construct the work PUMA level data.

Four dummies for the coastal work PUMAs of Atlantic Ocean, Pacific Ocean, Gulf of Mexico, and Great Lakes are constructed using the shapefile of the 2000 work PUMA. The coastline is also derived from the shapefile of the work PUMA to ensure consistency. If a work PUMA shares its boundary with any one of the four coastlines, it will be assigned a value of one for the corresponding coastal work PUMA dummy. Zeros are assigned to those work PUMAs not attached to any of the coastlines.

B. Full Results of Table 2

Appendix Table B.1 reports the full estimates of Table 2.

C. Agglomeration Wage Effect

The concept of urban wage premium is well documented in the literature: population geographic concentration increases wages and productivity (Ciccone, 2002; Rosenthal and Strange, 2004; Brülhart and Sbergami, 2009; Combes et al., 2010). To determine whether it works in the analytical sample of this paper, the tests are conducted as follows.

First, a reasonable measure of income should be constructed. Regression-adjusted hourly income is an appropriate option, which is computed as the work PUMA fixed effects from the model:

$$\ln(Hourly\ Income)_{icd} = \boldsymbol{X}_{icd}\boldsymbol{\beta} + \tau_d + \omega_c + \varepsilon_{icd} \tag{C.1}$$

where τ_d is the industry fixed effects. ω_c denotes the regression-adjusted average log hourly income in a work PUMA. Since people may have multiple sources of employment, the INCEARN in IPUMS is used as the annual income measure, which is the sum of the wage, business, and farm incomes in the previous year. Next, hourly incomes are obtained by dividing the annual incomes by the hours worked in the previous year for each observation. To obtain a more exogenous income measure, this regression is run for the self-employed and employed separately. Next, the regression-adjusted hourly income $\omega_c^{Self-Employed}$ and $\omega_c^{Employed}$ are obtained for the self-employed and employed, respectively. In the agglomeration wage effect test for the self-employed, $\omega_c^{Employed}$ is used as the welfare measure[23] because controlling for $\omega_c^{Employed}$ in the regressions could capture the spillover aspect of agglomeration, and could eliminate any mechanism endogeneity.

OLS is employed to estimate the agglomeration wage effect first. The regression model is similar to the preferred specification of the hours worked test:

$$\begin{aligned}\omega_c^{Employed} &= \alpha \log(Urbanization_c) + \beta \log(Localization_{cd}) + \boldsymbol{X}_{icd}\boldsymbol{\beta} \\ &\quad + \boldsymbol{A}_c\boldsymbol{\gamma} + \tau_d + \pi_m + \epsilon_{icd}\end{aligned} \tag{C.2}$$

This specification may suffer from endogeneity. Although one can include all the observable variables, unobserved variables might still exist. Additionally, the agglomeration measures may suffer from the measurement error bias. To ameliorate these issues and establish the causal relationship running from agglomeration to incomes, the same instruments are used as in the hours worked test, that is, the

[23] As a robustness check, $\omega_c^{Self-Employed}$ is also used as the alternative welfare measure.

minimum distance from the work PUMA centroid to the shoreline and estimated industry share in 1930.

Table C.1 reports the agglomeration wage effect by the OLS of Model (C.2). All estimates for urbanization are statistically significant and large in magnitude, implying that urbanization is correlated with higher wages. Although the estimates for localization are still positive, the magnitudes and significance are much lower than those for urbanization.

Considering the endogeneity issue, Table C.2 presents the 2SLS results for the agglomeration wage effect. The first row shows the 2SLS estimates for the log urbanization measure, which are not qualitatively different from the OLS estimates. However, when comparing the magnitudes, the effect of agglomeration on wages is understated by the OLS. The OLS results show that agglomeration increases wages by 0.0294–0.0315 log point for the different age groups of the lower educated self-employed, but 0.0638–0.0787 log point increases are estimated by the 2SLS. The agglomeration wage effects are approximately 0.0254–0.0264 for the highly educated self-employed by the OLS, but the estimates increase to 0.0423–0.0568 by the 2SLS. All the localization estimates decrease in magnitude and become insignificant. Most estimates even flip sign. Therefore, the agglomeration wage effect is from urbanization rather than localization, which confirms the literature.

Reference

Acs, Zoltan J., and Catherine Armington. 2006. *Entrepreneurship, Geography, and American Economic Growth*. New York: Cambridge University Press.

Angrist, Joshua D., and Guido W. Imbens. 1995. “Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity.” *Journal of the American Statistical Association* 90 (430): 431–42. doi:10.1080/01621459.1995.10476535.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. “Identification of Causal Effects Using Instrumental Variables.” *Journal of the American Statistical Association* 91 (434): 444–55. doi:10.1080/01621459.1996.10476902.

Aronson, Robert Louis. 1991. *Self-Employment: A Labour Market Perspective*. Ithaca, NY: ILR Press.

Bailey, Thomas, and Roger Waldinger. 1991. “Primary, Secondary, and Enclave Labor Markets: A Training Systems Approach.” *American Sociological Review* 56 (4): 432. doi:10.2307/2096266.

Baines, Susan, and Ulrike Gelder. 2003. “What Is Family Friendly about the Workplace in the Home? The Case of Self-Employed Parents and Their Children.” *New Technology, Work and Employment* 18 (3): 223–34. doi:10.1111/1468-005X.00123.

Baptista, Rui, Vítor Escária, and Paulo Madruga. 2008. “Entrepreneurship, Regional Development and Job Creation: The Case of Portugal.” *Small Business*

Economics 30 (1): 49–58. doi:10.1007/s11187-007-9055-0.

Bartik, Timothy J. 1991. *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, MI: W.E. Upjohn Institute for Employment Research. doi:10.17848/9780585223940.

Beeson, Patricia E., David N. DeJong, and Werner Troesken. 2001. “Population Growth in U.S. Counties, 1840-1990.” *Regional Science and Urban Economics* 31 (6): 669–99. doi:10.1016/S0166-0462(01)00065-5.

Blanchard, Olivier Jean, Lawrence F. Katz, Robert E. Hall, and Barry Eichengreen. 1992. “Regional Evolutions.” *Brookings Papers on Economic Activity* 1992 (1): 1–75. doi:10.2307/2534556.

Boden, Richard J. 1999. “Flexible Working Hours, Family Responsibilities, and Female Self-Employment.” *American Journal of Economics and Sociology* 58 (1): 71–83. doi:10.1111/j.1536-7150.1999.tb03285.x.

Brülhart, Marius, and Federica Sbergami. 2009. “Agglomeration and Growth: Cross-Country Evidence.” *Journal of Urban Economics* 65 (1): 48–63. doi:10.1016/j.jue.2008.08.003.

Cai, Zhengyu. 2018. “Why Does Spatial Equilibrium Lack for Strong Empirical Support? A Study Based on the Perspective of Researchers.” SWUFE Working Paper.

Carrington, William J, Kristin Mccue, and Brooks Pierce. 1996. “The Role of

Employer / Employee Interactions in Labor Market Cycles : Evidence from the Self-Employed." *Journal of Labor Economics* 14 (4): 571–602. doi:10.1086/209823.

Ciccone, Antonio. 2002. "Agglomeration Effects in Europe." *European Economic Review* 46 (2): 213–27. doi:10.1016/S0014-2921(00)00099-4.

Ciccone, Antonio, and Robert E. Hall. 1996. "Productivity and the Density of Economic Activity." *American Economic Review* 86 (1): 54–70. doi:10.1126/science.151.3712.867-a.

Clark, David E, William E Herrin, Thomas a Knapp, and Nancy E White. 2003. "Migration and Implicit Amenity Markets: Does Incomplete Compensation Matter? ." *Journal of Economic Geography* 3 (3): 289–307. doi:10.1093/jeg/3.3.289.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux. 2010. "Estimating Agglomeration Economies with History, Geology, and Worker Effects." In *Agglomeration Economics*, edited by Edward L. Glaeser, 15–66. Chicago, IL: University of Chicago Press.

Delgado, Mercedes, Michael E. Porter, and Scott Stern. 2010. "Clusters and Entrepreneurship." *Journal of Economic Geography* 10 (4): 495–518. doi:10.1093/jeg/lbq010.

Douglas, Evan J, and Dean A Shepherd. 2000. "Entrepreneurship as a Utility

Maximizing Response." *Journal of Business Venturing* 15 (3): 231–51. doi:10.1016/S0883-9026(98)00008-1.

Glaeser, Edward L., and David C. Maré. 2001. "Cities and Skills." *Journal of Labor Economics* 19 (2): 316–42. doi:10.1086/319563.

Glaeser, Edward L., William R. Kerr, and Giacomo A.M. Ponzetto. 2010. "Clusters of Entrepreneurship." *Journal of Urban Economics* 67 (1). Elsevier Inc.: 150–68. doi:10.1016/j.jue.2009.09.008.

Glaeser, Edward L., Stuart S. Rosenthal, and William C. Strange. 2010. "Urban Economics and Entrepreneurship." *Journal of Urban Economics* 67 (1). Elsevier Inc.: 1–14. doi:10.1016/j.jue.2009.10.005.

Greenwood, Michael J, G L Hunt, D S Rickman, and G I Treyz. 1991. "Migration, Regional Equilibrium, and the Estimation of Compensating Differentials." *American Economic Review* 81 (5): 382–90. doi:10.2307/2006927.

Gyourko, Joseph, Matthew Kahn, and Joseph Tracy. 1999. "Quality of Life and Environmental Comparisons." In *Handbook of Regional and Urban Economics*, edited by Edwin S Mills and Paul Cheshire, 3:1413–54. Elsevier. doi:10.1016/S1574-0080(99)80006-6.

Heckman, James J, Sergio Urzua, and Edward Vytlacil. 2006. "Understanding Instrumental Variables in Models with Essential Heterogeneity." *Review of Economics and Statistics* 88 (3): 389–432. doi:10.1162/rest.88.3.389.

Helsley, Robert W., and William C. Strange. 2011. “Entrepreneurs and Cities: Complexity, Thickness and Balance.” *Regional Science and Urban Economics* 41 (6): 550–59. doi:10.1016/j.regsciurbeco.2011.04.001.

Henderson, Jason, and Stephan Weiler. 2010. “Entrepreneurs and Job Growth: Probing the Boundaries of Time and Space.” *Economic Development Quarterly* 24 (1): 23–32. doi:10.1177/0891242409350917.

Imbens, G.W., and J.D. Angrist. 1994. “Identification and Estimation of Local Average Treatment Effects.” *Econometrica* 62 (2): 467–75. doi:10.2307/2951620.

Jarvis, A., H.I. I Reuter, A. Nelson, and E. Guevara. 2008. “Hole-Filled Seamless SRTM Data V4.” International Centre for Tropical Agriculture (CIAT). http://srtm.csi.cgiar.org.

Kahn, Matthew E. 2006. “Environmental Valuation Using Cross-City Hedonic Methods.” In *Environmental Valuation: Interregional and Intraregional Perspectives*, edited by John I. Carruthers and Bill Mundy, 27–48. Ashgate Publishing, Ltd.

Kahn, Shulamit, and Kevin Lang. 1991. “The Effect of Hours Constraints on Labor Supply Estimates.” *The Review of Economics and Statistics* 73 (4): 605–11. doi:10.2307/2109399.

Kleibergen, F, and R Paap. 2006. “Generalized Reduced Rank Tests Using the

Singular Value Decomposition.” *Journal of Econometrics* 133 (1): 97–126. doi:10.1016/j.jeconom.2005.02.011.

Krupka, Douglas J., and Kwame N. Donaldson. 2013. “Wages, Rents, and Heterogeneous Moving Costs.” *Economic Inquiry* 51 (1): 844–64. doi:10.1111/j.1465-7295.2012.00475.x.

Levine, Ross, and Yona Rubinstein. 2017. “Smart and Illicit: Who Becomes an Entrepreneur and Do They Earn More?” *Quarterly Journal of Economics* 132 (2): 963–1018. doi:10.1093/qje/qjw044.

Loscocco, Karyn A. 1997. “Work-Family Linkages among Self-Employed Women and Men.” *Journal of Vocational Behavior* 50 (2): 204–26. doi:10.1006/jvbe.1996.1576.

Moralee, Lisa. 1998. “Self-Employment in the 1990s.” *Labour Market Trends* 106 (3): 121–30.

Parker, Simon C. 2004. *The Economics of Self-Employment and Entrepreneurship*. New York: Cambridge University Press.

Parker, Simon C., Yacine Belghitar, and Tim Barmby. 2005. “Wage Uncertainty and the Labour Supply of Self-Employed Workers.” *Economic Journal* 115 (502): C190–207. doi:10.1111/j.0013-0133.2005.00987.x.

Portes, Alejandro, and Leif Jensen. 1992. “Disproving the Enclave Hypothesis: Reply.” *American Sociological Review* 57 (3): 418–20.

doi:10.1177/0003122412444721.

Portes, Alejandro, and Min Zhou. 1995. “Divergent Destinies: Immigration, Poverty, and Entrepreneurship.” In *Poverty, Inequality, and the Future of Social Policy,* edited by K. McFate, R. Lawson, and W. J. Wilson, 489–520. New York: Russell Sage Foundation.

———. 1996. “Self-Employment and the Earnings of Immigrants.” *American Sociological Review* 61 (2): 219–30. doi:10.2307/2096332.

Rappaport, Jordan, and Jeffrey D. Sachs. 2003. “The United States as a Coastal Nation.” *Journal of Economic Growth* 8 (1): 5–46. doi:10.1023/A:1022870216673.

Rees, Hedley, and Anup Shah. 1994. “The Characteristics of the Self-Employed: The Supply of Labour.” In *Employment, the Small Firm and the Labour Market,* edited by J. Atkinson and D.J. Storey, 317–27. London: Routledge.

Roback, Jennifer. 1988. “Wages, Rents, and Amenities: Differences Among Workers and Regions.” *Economic Inquiry* 26 (1): 23–41. doi:10.1111/j.1465-7295.1988.tb01667.x.

Rocha, Hector O. 2004. “Entrepreneurship and Development: The Role of Clusters.” *Small Business Economics*. doi:10.1007/s11187-004-3991-8.

Rosenthal, Stuart S., and William C. Strange. 2008. “Agglomeration and Hours Worked.” *Review of Economics and Statistics* 90 (1): 105–18.

doi:10.1162/rest.90.1.105.

Rosenthal, Stuart S, and William C Strange. 2004. “Evidence on the Nature and Sources of Agglomeration Economies.” In *Handbook of Regional and Urban Economics*, edited by Vernon Henderson and Jacques-François Thisse, 4:2119–71. Amsterdam: North-Holland. doi:10.1016/S0169-7218(04)07049-2.

Ruggles, S, J T Alexander, K Genadek, R Goeken, M B Schroeder, and M Sobek. 2010. “Integrated Public Use Microdata Series: Version 5.0 [Machine-Readable Database].” Minneapolis: University of Minnesota. doi:http://doi.org/10.18128/D010.V6.0.

Small, Christopher, and Joel E. Cohen. 2004. “Continental Physiography, Climate, and the Global Distribution of Human Population.” *Current Anthropology* 45 (2): 269–77.

Small, Christopher, Vivien Gornitz, and Joel E. Cohen. 2000. “Coastal Hazards and Global Distribution of Human Population.” *Environmental Geosciences* 7 (1): 3–12. doi:10.1046/j.1526-0984.2000.71005.x.

Small, Christopher, and Robert J Nicholls. 2003. “A Global Analysis of Human Settlement in Coastal Zones.” *Journal of Coastal Research* 19 (3): 584–99. doi:10.2307/4299200.

Stephens, Heather M., and Mark D. Partridge. 2011. “Do Entrepreneurs Enhance Economic Growth in Lagging Regions?” *Growth and Change* 42 (4): 431–65.

doi:10.1111/j.1468-2257.2011.00563.x.

Stephens, Heather M., Mark D. Partridge, and Alessandra Faggian. 2013. “Innovation, Entrepreneurship and Economic Growth in Lagging Regions.” *Journal of Regional Science* 53 (5): 778–812. doi:10.1111/jors.12019.

Stock, James H., and Motohiro Yogo. 2005. “Testing for Weak Instruments in Linear IV Regression.” In *Identification and Inference for Econometric Models*, edited by Donald W. K. Andrews and James H. Stock, 80–108. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511614491.006.

“United States Department of Justice. Federal Bureau of Investigation. Uniform Crime Reporting Program Data [United States]: County-Level Detailed Arrest and Offense Data.” 2000. ICPSR03451-v4. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2006-01-16. doi:10.3886/ICPSR03451.v4.

Winters, John V. 2013. “Human Capital Externalities and Employment Differences across Metropolitan Areas of the USA.” *Journal of Economic Geography* 13 (5): 799–822. doi:10.1093/jeg/lbs046.

Yankow, Jeffrey J. 2006. “Why Do Cities Pay More? An Empirical Examination of Some Competing Theories of the Urban Wage Premium.” *Journal of Urban Economics* 60 (2): 139–61. doi:10.1016/j.jue.2006.03.004.

Table 1: Summary Statistics for Analytical Sample

	No. Obs.	Mean	Std. Dev.	Min	Max
Hours worked	158345	50.877	12.063	35.000	99.000
Urbanization	158345	3133.976	10585.298	1.637	66942.258
Localization	158345	8.947	53.189	0.000	716.979
Log (minimum distance to coastline)	158345	3.934	1.818	-3.772	6.706
Log (industry share in 1930)	158345	-4.114	1.726	-11.259	0.000
High school and less	158345	0.531	0.499	0.000	1.000
College and more	158345	0.469	0.499	0.000	1.000
Age	158345	44.759	7.854	30.000	59.000
Commute time	158345	22.307	27.125	0.000	174.000
Children present	158345	0.594	0.491	0.000	1.000
Education					
No schooling completed	158345	0.006	0.076	0.000	1.000
Nursery school to grade 4	158345	0.004	0.060	0.000	1.000
Grade 5 or 6	158345	0.011	0.104	0.000	1.000
Grade 7 or 8	158345	0.023	0.149	0.000	1.000
Grade 9	158345	0.020	0.141	0.000	1.000
Grade 10	158345	0.029	0.167	0.000	1.000
Grade 11	158345	0.027	0.163	0.000	1.000
Grade 12, no diploma	158345	0.040	0.197	0.000	1.000
High school graduate or GED	158345	0.371	0.483	0.000	1.000
Bachelor's degree	158345	0.257	0.437	0.000	1.000
Master's degree	158345	0.070	0.255	0.000	1.000
Professional degree beyond a bachelor's	158345	0.123	0.329	0.000	1.000
Doctoral degree	158345	0.019	0.137	0.000	1.000
Marital Status					
Married	158345	0.778	0.416	0.000	1.000
Married, spouse absent	158345	0.011	0.102	0.000	1.000
Separated	158345	0.014	0.117	0.000	1.000
Divorced	158345	0.100	0.299	0.000	1.000
Widowed	158345	0.005	0.073	0.000	1.000
Never married	158345	0.093	0.290	0.000	1.000
Race					
White	158345	0.880	0.325	0.000	1.000
African American	158345	0.035	0.183	0.000	1.000
American Indian or Alaska Native	158345	0.004	0.066	0.000	1.000
Chinese	158345	0.011	0.104	0.000	1.000
Japanese	158345	0.002	0.047	0.000	1.000
Other Asian or Pacific Islander	158345	0.026	0.160	0.000	1.000
Other race	158345	0.025	0.157	0.000	1.000

Two major races	158345	0.015	0.123	0.000	1.000
Three or more major races	158345	0.001	0.024	0.000	1.000
Hispanic Origin					
Not Hispanic	158345	0.936	0.245	0.000	1.000
Mexican	158345	0.036	0.186	0.000	1.000
Puerto Rican	158345	0.003	0.056	0.000	1.000
Cuban	158345	0.005	0.074	0.000	1.000
Other	158345	0.019	0.137	0.000	1.000
Years of Residency in the USA					
Native	158345	0.864	0.343	0.000	1.000
Years in USA 0-5	158345	0.010	0.098	0.000	1.000
Years in USA 6-10	158345	0.016	0.126	0.000	1.000
Years in USA 11-15	158345	0.023	0.149	0.000	1.000
Years in USA 16-20	158345	0.025	0.156	0.000	1.000
Years in USA 20+	158345	0.063	0.242	0.000	1.000
Amenities					
Log (violent crime)	158102	6.635	1.703	1.946	10.380
Log (property crime)	158167	7.824	1.522	2.639	12.000
Log (precipitation)	158345	8.950	0.459	6.677	9.998
Log (dew points)	158345	7.148	0.727	-10.735	7.855
Log (January temperature)	158345	2.737	0.647	-13.356	3.540
Log (July temperature)	158345	3.176	0.142	2.608	3.499
Log (heating degree days)	158345	7.540	0.864	3.664	8.616
Log (cooling degree days)	158345	7.660	0.430	6.157	8.556
Log (elevation)	158345	8.406	0.811	-7.953	8.904
Log (solar irradiance)	158345	1.512	0.201	1.129	2.069
Log (minimum distance to river and lake)	158345	2.145	1.199	-2.052	4.432
Atlantic work PUMA	158345	0.155	0.362	0.000	1.000
Great Lake work PUMA	158345	0.053	0.224	0.000	1.000
Gulf work PUMA	158345	0.052	0.223	0.000	1.000
Pacific work PUMA	158345	0.120	0.325	0.000	1.000

Note: All summary statistics are adjusted by personal weight to ensure the national representative.
Industry, work PUMA, and work MSA are not included for space conservation.

Table 2: Hours Worked and Agglomeration Using OLS

	(1)	(2)	(3)	(4)
Log (urbanization)	-0.0176***	-0.0184***	-0.0180***	-0.0187***
	(0.0016)	(0.0017)	(0.0016)	(0.0017)
Log (localization)	0.0120***	0.0122***	0.0122***	0.0123***
	(0.0010)	(0.0010)	(0.0010)	(0.0010)
Self-employed share at the work PUMA level		-0.0637		-0.0548
		(0.0428)		(0.0431)
Self-employed share within industries at the work PUMA level			-0.0098*	-0.0090
			(0.0057)	(0.0058)
N	137,621	137,621	137,621	137,621
R^2	0.068	0.068	0.068	0.068

Notes: Dependent variable is log usual hours worked per week in the previous year. Regressions include all the controls listed in Table 1. Industry dummies and work MSA dummies are also included. Other estimates are suppressed for space conservation and are available in Appendix Table B.1. Standard errors in parentheses are robust to heteroskedasticity and clustered by work PUMA. * $p < 0.1$, *** $p < 0.01$.

Table 3: Hours Worked and Agglomeration by Age and Education Using OLS

	High school and less			College and more		
	(1)	(2)	(3)	(4)	(5)	(6)
	Age 30-39	Age 40-49	Age 50-59	Age 30-39	Age 40-49	Age 50-59
Log (urbanization)	-0.0281***	-0.0199***	-0.0173***	-0.0087*	-0.0075**	-0.0191***
	(0.0033)	(0.0032)	(0.0035)	(0.0045)	(0.0035)	(0.0036)
Log (localization)	0.0223***	0.0118***	0.0121***	0.0078***	0.0076***	0.0104***
	(0.0028)	(0.0023)	(0.0023)	(0.0023)	(0.0018)	(0.0019)
N	25,813	31,501	21,659	13,724	23,807	21,117
R^2	0.113	0.110	0.111	0.068	0.047	0.054

Notes: Dependent variable is log usual hours worked per week in the previous year. Regressions include all the controls listed in Table 1. Industry dummies and work MSA dummies are also included. Other estimates are suppressed for space conservation. Standard errors in parentheses are robust to heteroskedasticity and clustered by work PUMA. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Hours Worked and Agglomeration Using 2SLS

	High school and less			College and more		
	(1)	(2)	(3)	(4)	(5)	(6)
	Age 30-39	Age 40-49	Age 50-59	Age 30-39	Age 40-49	Age 50-59
Log (urbanization)	-0.0477***	-0.0478***	-0.0694***	-0.0173	-0.0439***	-0.0451***
	(0.0167)	(0.0118)	(0.0178)	(0.0160)	(0.0136)	(0.0115)
Log (localization)	0.0388***	0.0392***	0.0585***	0.0199***	0.0262***	0.0271***
	(0.0072)	(0.0067)	(0.0083)	(0.0075)	(0.0055)	(0.0058)
First Stage: Log (urbanization)						
Log (distance to shoreline)	-0.2830***	-0.2971***	-0.2671***	-0.5015***	-0.4745***	-0.4564***
	(0.0578)	(0.0558)	(0.0545)	(0.0664)	(0.0613)	(0.0610)
Log (industry share in 1930)	0.0407***	0.0487***	0.0405***	0.0591***	0.0559***	0.0471***
	(0.0108)	(0.0097)	(0.0093)	(0.0109)	(0.0100)	(0.0097)
First Stage: Log (localization)						
Log (distance to shoreline)	-0.3098***	-0.3107***	-0.3072***	-0.7220***	-0.6672***	-0.6166***
	(0.0692)	(0.0664)	(0.0686)	(0.1310)	(0.1126)	(0.0986)
Log (industry share in 1930)	0.3153***	0.3307***	0.3105***	0.4069***	0.4008***	0.3803***
	(0.0178)	(0.0170)	(0.0162)	(0.0205)	(0.0200)	(0.0207)
Underidentification	18.8372	25.1217	23.3020	39.1858	42.0089	41.3153
	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
Weak identification	13.0326	16.3249	14.5964	42.1833	45.6666	45.1562
Endogeneity	6.1387	24.3959	34.1100	6.0584	11.9176	8.9949
	[0.0465]	[0.0000]	[0.0000]	[0.0484]	[0.0026]	[0.0111]

Notes: Dependent variable is log usual hours worked per week in the previous year. Regressions include all the controls listed in Table 1. Industry dummies and work MSA dummies are also included. Other estimates are suppressed for space conservation. Standard errors in parentheses are robust to heteroskedasticity and clustered by work PUMA. P-values are provided in square brackets for underidentification tests and endogeneity tests. *** $p < 0.01$.

Table 5: Robustness Checks

	High school and less			College and more		
	(1)	(2)	(3)	(4)	(5)	(6)
	Age 30-39	Age 40-49	Age 50-59	Age 30-39	Age 40-49	Age 50-59
A. Exclusion of Port Work PUMAs						
Log (urbanization)	-0.0672**	-0.0353*	-0.0627**	-0.0200	-0.0386**	-0.0313*
	(0.0304)	(0.0201)	(0.0296)	(0.0246)	(0.0191)	(0.0170)
Log (localization)	0.0385***	0.0377***	0.0551***	0.0213**	0.0271***	0.0259***
	(0.0092)	(0.0078)	(0.0102)	(0.0095)	(0.0067)	(0.0067)
B. Extra Instrument						
Log (urbanization)	-0.0454***	-0.0479***	-0.0646***	-0.0255***	-0.0289***	-0.0336***
	(0.0078)	(0.0071)	(0.0093)	(0.0094)	(0.0069)	(0.0072)
Log (localization)	0.0384***	0.0392***	0.0577***	0.0218***	0.0229***	0.0249***
	(0.0067)	(0.0066)	(0.0076)	(0.0069)	(0.0049)	(0.0055)
Hansen J overidentification	0.0325	0.0001	0.1136	0.4428	1.5519	1.5424
	[0.8569]	[0.9933]	[0.7361]	[0.5058]	[0.2129]	[0.2143]
C. Control for Bartik Shift-Share						
Log (urbanization)	-0.0455**	-0.0501***	-0.0821***	-0.0062	-0.0536***	-0.0530***
	(0.0212)	(0.0153)	(0.0200)	(0.0184)	(0.0136)	(0.0146)
Log (localization)	0.0342***	0.0348***	0.0509***	0.0148*	0.0232***	0.0280***
	(0.0107)	(0.0113)	(0.0119)	(0.0083)	(0.0063)	(0.0072)
Bartik Shift-Share	-0.5854**	-0.3428	-0.2652	-0.3360	-0.2178	-0.3465
	(0.2357)	(0.2100)	(0.2181)	(0.2810)	(0.2199)	(0.2329)
D. Relative Localization						
Log (urbanization)	-0.0647**	-0.0817***	-0.1025***	-0.0123	-0.0404***	-0.0544***
	(0.0260)	(0.0313)	(0.0376)	(0.0153)	(0.0145)	(0.0176)
Log (localization)/Log (urbanization)	0.2418***	0.2356***	0.3388***	0.1050***	0.1331***	0.1525***
	(0.0483)	(0.0486)	(0.0563)	(0.0396)	(0.0282)	(0.0331)
E. Work PUMA Fixed Effects						
Log (localization)	0.0381***	0.0345***	0.0562***	0.0208***	0.0237***	0.0261***
	(0.0064)	(0.0063)	(0.0074)	(0.0066)	(0.0047)	(0.0054)

F. Cluster by State						
Log (urbanization)	-0.0477***	-0.0478***	-0.0694***	-0.0173	-0.0439**	-0.0451***
	(0.0169)	(0.0107)	(0.0150)	(0.0141)	(0.0209)	(0.0173)
Log (localization)	0.0388***	0.0392***	0.0585***	0.0199***	0.0262***	0.0271***
	(0.0055)	(0.0060)	(0.0070)	(0.0073)	(0.0080)	(0.0066)
G. Immigrants						
Log (urbanization)	-0.1108***	-0.0523	-0.0966***	-0.0942**	-0.0586*	-0.0419
	(0.0373)	(0.0344)	(0.0368)	(0.0380)	(0.0319)	(0.0339)
Log (localization)	0.0508***	0.0382*	0.0628***	0.0450**	0.0330**	0.0313**
	(0.0159)	(0.0224)	(0.0188)	(0.0193)	(0.0131)	(0.0159)
H. Natives						
Log (urbanization)	-0.0270	-0.0408***	-0.0475**	-0.0041	-0.0430***	-0.0496***
	(0.0168)	(0.0145)	(0.0204)	(0.0191)	(0.0152)	(0.0128)
Log (localization)	0.0271***	0.0345***	0.0526***	0.0131	0.0246***	0.0289***
	(0.0081)	(0.0075)	(0.0091)	(0.0082)	(0.0060)	(0.0061)

Notes: All regressions are estimated by 2SLS. Dependent variable is log usual hours worked per week in the previous year. Regressions include all the controls listed in Table 1. Industry dummies and work MSA dummies are also included. Other estimates and the first stage estimates are suppressed for space conservation. Standard errors in parentheses are robust to heteroskedasticity and clustered by work PUMA. P-values are provided in square brackets for overidentification tests. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Hours Worked and Competition Using OLS

	High school and less			College and more		
	(1)	(2)	(3)	(4)	(5)	(6)
	Age 30-39	Age 40-49	Age 50-59	Age 30-39	Age 40-49	Age 50-59
Not Incorporated						
Peer pressure	0.0137***	0.0064	0.0037	0.0202***	0.0084**	0.0036
	(0.0042)	(0.0040)	(0.0044)	(0.0064)	(0.0041)	(0.0048)
Large firm pressure	0.0072**	0.0072**	0.0106***	-0.0002	0.0079**	0.0089**
	(0.0036)	(0.0031)	(0.0035)	(0.0061)	(0.0037)	(0.0045)
Log (urbanization)	-0.0062*	-0.0079***	-0.0077**	-0.0079	-0.0006	-0.0148***
	(0.0035)	(0.0030)	(0.0038)	(0.0054)	(0.0043)	(0.0045)
Incorporated						
Peer pressure	0.0053	0.0022	-0.0079	0.0093*	-0.0016	0.0084*
	(0.0063)	(0.0047)	(0.0057)	(0.0053)	(0.0041)	(0.0047)
Small firm pressure	0.0027	0.0072	0.0056	0.0006	0.0104***	0.0045
	(0.0059)	(0.0044)	(0.0053)	(0.0050)	(0.0038)	(0.0038)
Log (urbanization)	-0.0125**	-0.0106***	-0.0002	0.0003	-0.0063	-0.0061
	(0.0058)	(0.0041)	(0.0051)	(0.0054)	(0.0042)	(0.0044)

Notes: Dependent variable is log usual hours worked per week in the previous year. Regressions include all the controls listed in Table 1. Industry dummies and work MSA dummies are also included. Other estimates are suppressed for space conservation. Standard errors in parentheses are robust to heteroskedasticity and clustered by work PUMA. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix Table B.1: Hours Worked and Agglomeration Using OLS

	(1)	(2)	(3)	(4)
Log (urbanization)	-0.0176***	-0.0184***	-0.0180***	-0.0187***
	(0.0016)	(0.0017)	(0.0016)	(0.0017)
Log (localization)	0.0120***	0.0122***	0.0122***	0.0123***
	(0.0010)	(0.0010)	(0.0010)	(0.0010)
Self-employed share at the work PUMA level		-0.0637		-0.0548
		(0.0428)		(0.0431)
Self-employed share of industries at the work PUMA level			-0.0098*	-0.0090
			(0.0057)	(0.0058)
Age	0.0139	0.0139	0.0141	0.0141
	(0.0596)	(0.0596)	(0.0596)	(0.0596)
Age^2	-0.0006	-0.0006	-0.0006	-0.0006
	(0.0020)	(0.0020)	(0.0020)	(0.0020)
Age^3	0.0000	0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Age^4	-0.0000	-0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Children present	0.0060***	0.0060***	0.0060***	0.0060***
	(0.0016)	(0.0016)	(0.0016)	(0.0016)
Log (commute time)	-0.0076***	-0.0076***	-0.0076***	-0.0076***
	(0.0008)	(0.0008)	(0.0008)	(0.0008)
Log (violent crime)	0.0037	0.0035	0.0037	0.0036
	(0.0023)	(0.0023)	(0.0023)	(0.0023)
Log (property crime)	-0.0030	-0.0034	-0.0032	-0.0035
	(0.0028)	(0.0028)	(0.0028)	(0.0028)
Log (precipitation)	-0.0163**	-0.0150**	-0.0159**	-0.0148**
	(0.0070)	(0.0071)	(0.0070)	(0.0071)
Log (January temperature)	0.0010	0.0011	0.0010	0.0011
	(0.0008)	(0.0008)	(0.0008)	(0.0008)

Log (July temperature)	0.0790*	0.0777*	0.0789*	0.0778*
	(0.0437)	(0.0441)	(0.0437)	(0.0440)
Log (elevation)	-0.0012***	-0.0012***	-0.0012***	-0.0011***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Log (minimum distance to river and lake)	0.0018*	0.0018*	0.0018*	0.0018*
	(0.0010)	(0.0010)	(0.0010)	(0.0010)
Log (heating degree days)	0.0018	0.0012	0.0018	0.0012
	(0.0064)	(0.0065)	(0.0064)	(0.0065)
Log (cooling degree days)	-0.0253	-0.0262	-0.0253	-0.0261
	(0.0205)	(0.0206)	(0.0205)	(0.0206)
Log (dew points)	0.0011*	0.0010	0.0011*	0.0011*
	(0.0006)	(0.0007)	(0.0006)	(0.0007)
Log (solar irradiance)	-0.0015	0.0019	-0.0018	0.0012
	(0.0268)	(0.0268)	(0.0268)	(0.0267)
Atlantic work PUMA	-0.0012	-0.0006	-0.0010	-0.0005
	(0.0046)	(0.0046)	(0.0046)	(0.0046)
Great Lake work PUMA	0.0015	0.0021	0.0018	0.0023
	(0.0061)	(0.0063)	(0.0062)	(0.0063)
Gulf work PUMA	-0.0134	-0.0128	-0.0133	-0.0128
	(0.0084)	(0.0086)	(0.0085)	(0.0086)
Pacific work PUMA	0.0291***	0.0294***	0.0292***	0.0294***
	(0.0108)	(0.0108)	(0.0108)	(0.0109)
Education: reference category - No schooling completed				
Nursery school to grade 4	-0.0273**	-0.0275**	-0.0274**	-0.0275**
	(0.0138)	(0.0138)	(0.0139)	(0.0139)
Grade 5 or 6	-0.0147	-0.0148	-0.0147	-0.0147
	(0.0104)	(0.0104)	(0.0104)	(0.0104)
Grade 7 or 8	-0.0121	-0.0122	-0.0119	-0.0120
	(0.0098)	(0.0098)	(0.0098)	(0.0098)

Grade 9	-0.0044	-0.0044	-0.0042	-0.0043
	(0.0101)	(0.0101)	(0.0101)	(0.0101)
Grade 10	-0.0092	-0.0093	-0.0091	-0.0092
	(0.0099)	(0.0099)	(0.0099)	(0.0099)
Grade 11	-0.0063	-0.0063	-0.0062	-0.0063
	(0.0099)	(0.0099)	(0.0099)	(0.0099)
Grade 12, no diploma	0.0006	0.0006	0.0007	0.0007
	(0.0100)	(0.0100)	(0.0100)	(0.0100)
High school graduate or GED	0.0134	0.0134	0.0135	0.0135
	(0.0093)	(0.0093)	(0.0093)	(0.0093)
Bachelor's degree	0.0254***	0.0254***	0.0254***	0.0254***
	(0.0092)	(0.0092)	(0.0092)	(0.0092)
Master's degree	0.0288***	0.0288***	0.0286***	0.0286***
	(0.0093)	(0.0093)	(0.0093)	(0.0093)
Professional degree beyond a bachelor's	0.0356***	0.0355***	0.0354***	0.0353***
	(0.0092)	(0.0092)	(0.0092)	(0.0092)
Doctoral degree	0.0250**	0.0249**	0.0245**	0.0245**
	(0.0100)	(0.0100)	(0.0100)	(0.0100)
Marital status: reference category - Married				
Married, spouse absent	-0.0128*	-0.0127*	-0.0127*	-0.0127*
	(0.0074)	(0.0074)	(0.0074)	(0.0074)
Separated	-0.0237***	-0.0237***	-0.0236***	-0.0237***
	(0.0050)	(0.0050)	(0.0050)	(0.0050)
Divorced	-0.0265***	-0.0265***	-0.0265***	-0.0265***
	(0.0023)	(0.0023)	(0.0023)	(0.0023)
Widowed	-0.0405***	-0.0406***	-0.0404***	-0.0405***
	(0.0080)	(0.0080)	(0.0080)	(0.0080)
Never married	-0.0359***	-0.0359***	-0.0359***	-0.0359***
	(0.0025)	(0.0025)	(0.0025)	(0.0025)

Race: reference category - White				
African American	-0.0189***	-0.0190***	-0.0189***	-0.0189***
	(0.0045)	(0.0045)	(0.0045)	(0.0045)
American Indian or Alaska Native	0.0098	0.0096	0.0097	0.0096
	(0.0097)	(0.0097)	(0.0097)	(0.0097)
Chinese	0.0103	0.0102	0.0101	0.0101
	(0.0134)	(0.0134)	(0.0134)	(0.0134)
Japanese	-0.0205	-0.0206	-0.0206	-0.0207
	(0.0143)	(0.0143)	(0.0144)	(0.0143)
Other Asian or Pacific Islander	0.0130**	0.0129**	0.0129**	0.0128**
	(0.0057)	(0.0057)	(0.0056)	(0.0057)
Other race	-0.0105*	-0.0105*	-0.0105*	-0.0105*
	(0.0058)	(0.0058)	(0.0058)	(0.0058)
Two major races	0.0140**	0.0140**	0.0141**	0.0141**
	(0.0062)	(0.0062)	(0.0062)	(0.0062)
Three or more major races	0.0673*	0.0673*	0.0674*	0.0674*
	(0.0386)	(0.0386)	(0.0386)	(0.0386)
Hispanic origin: reference category - Not Hispanic				
Mexican	-0.0474***	-0.0475***	-0.0475***	-0.0476***
	(0.0044)	(0.0044)	(0.0044)	(0.0044)
Puerto Rican	-0.0078	-0.0078	-0.0079	-0.0079
	(0.0128)	(0.0128)	(0.0128)	(0.0128)
Cuban	-0.0021	-0.0021	-0.0021	-0.0020
	(0.0134)	(0.0134)	(0.0134)	(0.0134)
Other	-0.0288***	-0.0288***	-0.0288***	-0.0288***
	(0.0054)	(0.0054)	(0.0054)	(0.0054)
Years of residency in the US: reference category - Years in USA 0-5				
Native	0.0316***	0.0315***	0.0316***	0.0315***
	(0.0068)	(0.0068)	(0.0068)	(0.0068)

Years in USA 6-10	0.0395***	0.0394***	0.0395***	0.0395***
	(0.0088)	(0.0088)	(0.0088)	(0.0088)
Years in USA 11-15	0.0447***	0.0447***	0.0447***	0.0447***
	(0.0089)	(0.0089)	(0.0089)	(0.0089)
Years in USA 16-20	0.0512***	0.0512***	0.0513***	0.0513***
	(0.0078)	(0.0078)	(0.0078)	(0.0078)
Years in USA 20+	0.0568***	0.0568***	0.0567***	0.0567***
	(0.0069)	(0.0069)	(0.0069)	(0.0069)
Constant	3.9918***	4.0062***	3.9950***	4.0071***
	(0.6747)	(0.6753)	(0.6752)	(0.6757)
N	137,621	137,621	137,621	137,621
R^2	0.068	0.068	0.068	0.068

Notes: Dependent variable is log usual hours worked per week in the previous year. Industry dummies and work MSA dummies are suppressed for space conservation. Standard errors in parentheses are robust to heteroskedasticity and clustered by work PUMA. * p < 0.1, ** p < 0.05, *** p < 0.01.

Appendix Table C.1: Agglomeration Wage Effect Using OLS

	High school and less			College and more		
	(1)	(2)	(3)	(4)	(5)	(6)
	Age 30-39	Age 40-49	Age 50-59	Age 30-39	Age 40-49	Age 50-59
Log(urbanization)	0.0294***	0.0315***	0.0308***	0.0264***	0.0259***	0.0254***
	(0.0031)	(0.0031)	(0.0032)	(0.0037)	(0.0036)	(0.0036)
Log(localization)	0.0011	0.0009	0.0003	0.0042*	0.0039*	0.0037**
	(0.0017)	(0.0015)	(0.0014)	(0.0022)	(0.0020)	(0.0016)
N	25,813	31,501	21,659	13,724	23,807	21,117
R^2	0.899	0.896	0.895	0.935	0.932	0.932

Notes: Dependent variable is regression-adjusted average log hourly income of paid workers. Regressions include all the controls listed in Table 1. Industry dummies and work MSA dummies are also included. Other estimates are suppressed for space conservation. Standard errors in parentheses are robust to heteroskedasticity and clustered by work PUMA. * p < 0.1, ** p < 0.05, *** p < 0.01.

Appendix Table C.2: Agglomeration Wage Effect Using 2SLS

	High school and less			College and more		
	(1)	(2)	(3)	(4)	(5)	(6)
	Age 30-39	Age 40-49	Age 50-59	Age 30-39	Age 40-49	Age 50-59
Log (urbanization)	0.0638***	0.0651***	0.0787***	0.0568***	0.0495***	0.0423***
	(0.0203)	(0.0185)	(0.0228)	(0.0203)	(0.0184)	(0.0159)
Log (localization)	-0.0004	-0.0014	-0.0050	-0.0013	-0.0006	0.0023
	(0.0034)	(0.0032)	(0.0037)	(0.0029)	(0.0028)	(0.0022)
First Stage: Log (urbanization)						
Log (distance to shoreline)	-0.2830***	-0.2971***	-0.2671***	-0.5015***	-0.4745***	-0.4564***
	(0.0578)	(0.0558)	(0.0545)	(0.0664)	(0.0613)	(0.0610)
Log (industry share in 1930)	0.0407***	0.0487***	0.0405***	0.0591***	0.0559***	0.0471***
	(0.0108)	(0.0097)	(0.0093)	(0.0109)	(0.0100)	(0.0097)
First Stage: Log (localization)						
Log (distance to shoreline)	-0.3098***	-0.3107***	-0.3072***	-0.7220***	-0.6672***	-0.6166***
	(0.0692)	(0.0664)	(0.0686)	(0.1310)	(0.1126)	(0.0986)
Log (industry share in 1930)	0.3153***	0.3307***	0.3105***	0.4069***	0.4008***	0.3803***
	(0.0178)	(0.0170)	(0.0162)	(0.0205)	(0.0200)	(0.0207)
Underidentification	18.8372	25.1217	23.3020	39.1858	42.0089	41.3153
	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
Weak identification	13.0326	16.3249	14.5964	42.1833	45.6666	45.1562
Endogeneity	11.5340	13.1315	13.8964	5.4750	3.1569	4.9515
	[0.0031]	[0.0014]	[0.0010]	[0.0647]	[0.2063]	[0.0841]

Notes: Dependent variable is regression-adjusted average log hourly income of paid workers. Regressions include all the controls listed in Table 1. Industry dummies and work MSA dummies are also included. Other estimates are suppressed for space conservation. Standard errors in parentheses are robust to heteroskedasticity and clustered by work PUMA. P-values are provided in square brackets for underidentification tests and endogeneity tests. *** $p < 0.01$.

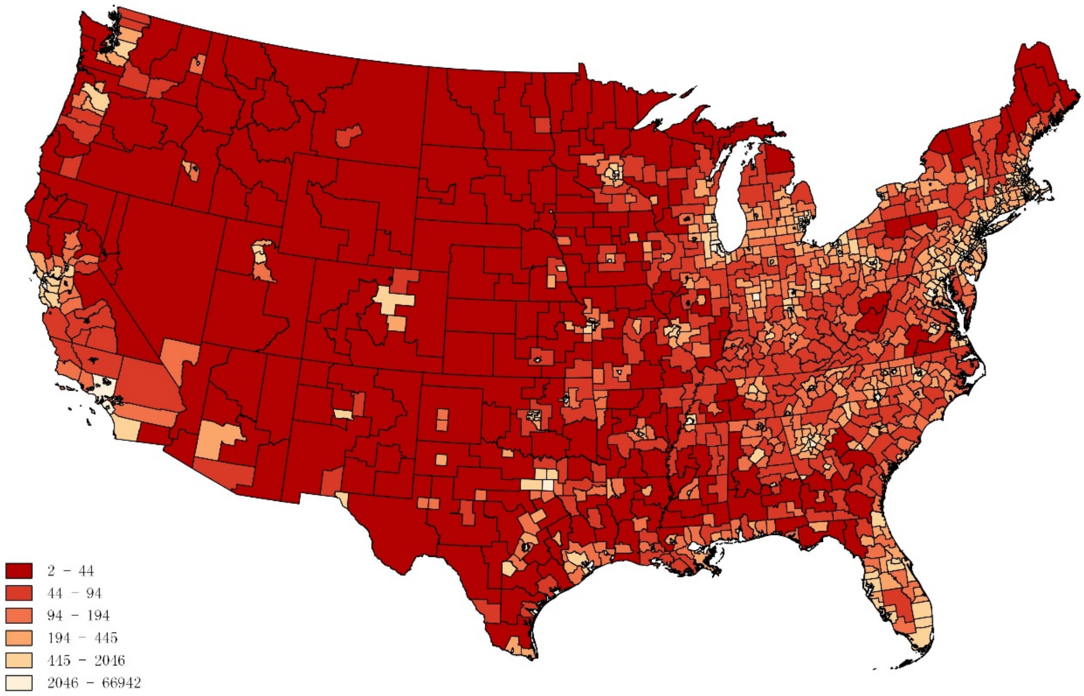

Figure 1: 2000 Population density at work PUMA level. (The magnitudes are classified by quantiles.) Source: Author.

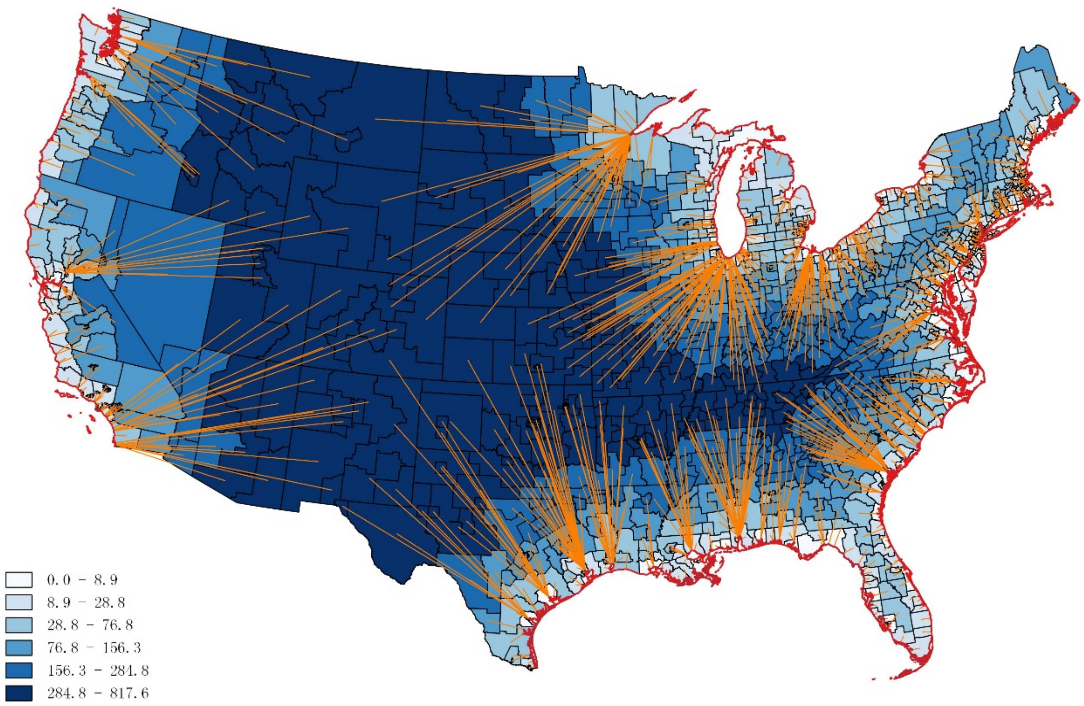

Figure 2: Minimum distance from work PUMA centroid to coastline. (The magnitudes in miles are classified by quantiles.) Source: Author.

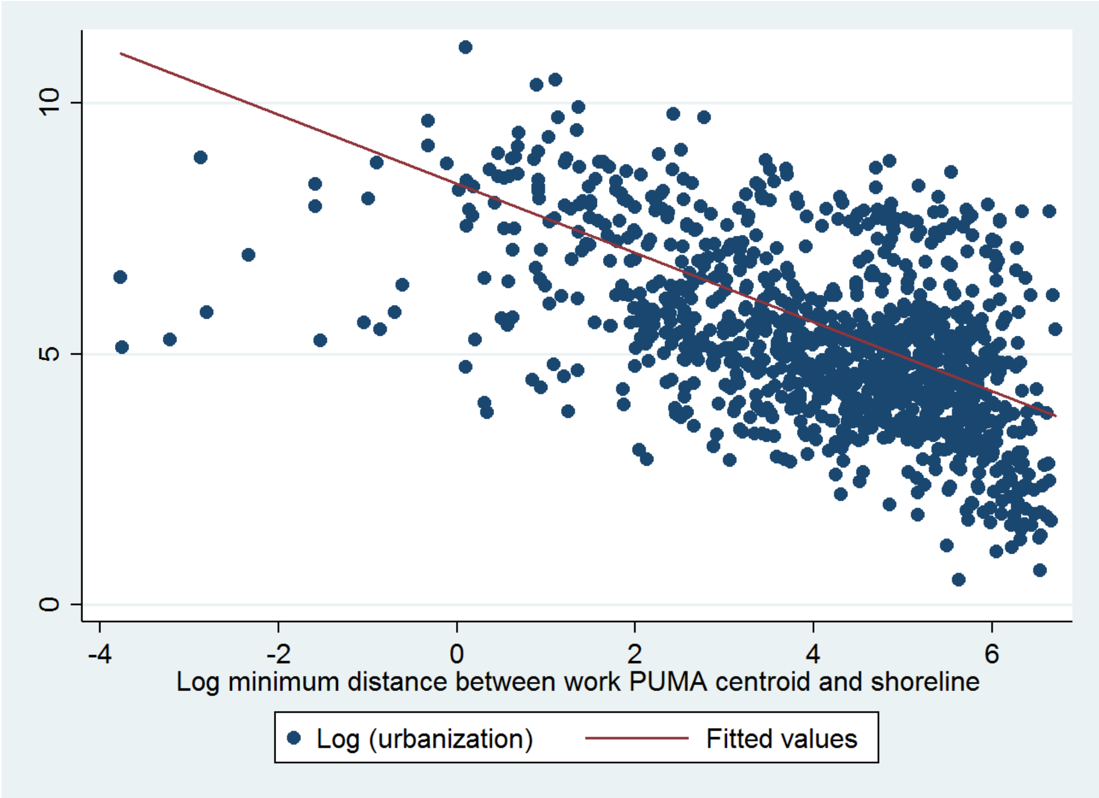

Figure 3: Raw correlation between log urbanization and log minimum distance to shoreline. Source: Author.

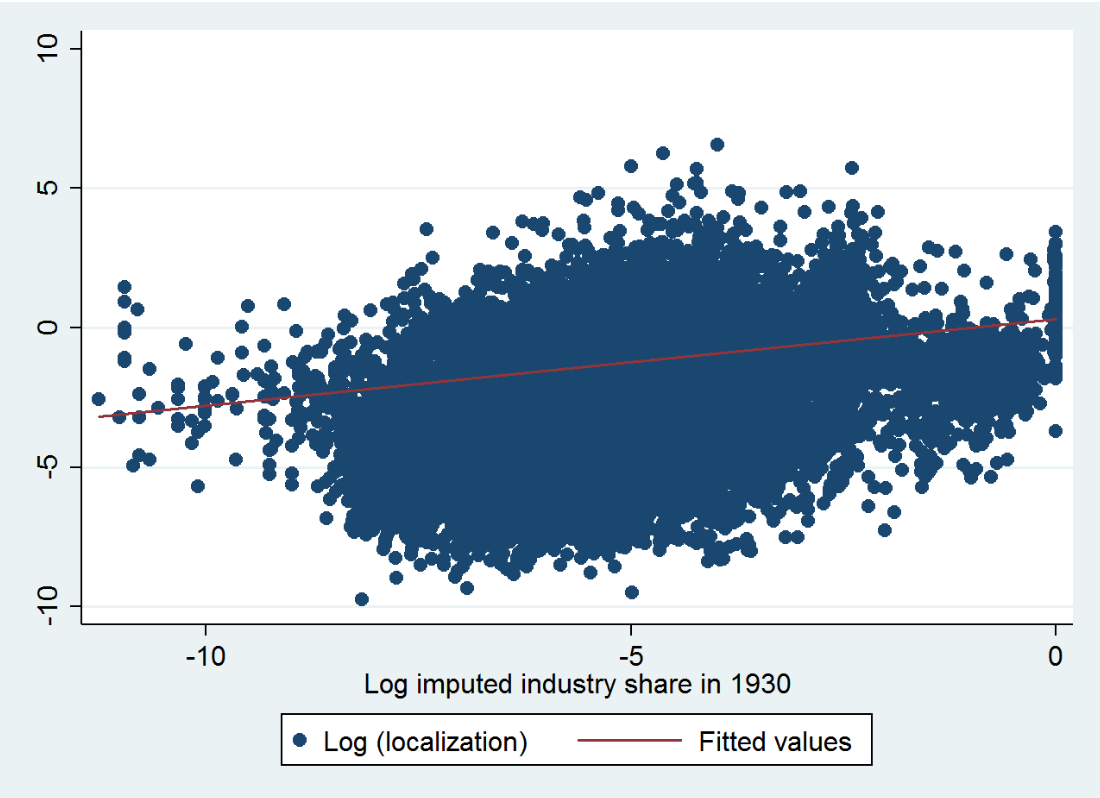

Figure 4: Raw correlation between log localization and log industry share in 1930. Source: Author.