Online Appendix

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:

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

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[1] because controlling for

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

$\omega_c^{Employed}$ 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:

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

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 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 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.

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. *** $p < 0.01$.