

IZA DP No. 4573

## **Some Evidence on the Nature of Urbanization Economies**

Douglas J. Krupka

November 2009

# Some Evidence on the Nature of Urbanization Economies

**Douglas J. Krupka**

*University of Michigan  
and IZA*

Discussion Paper No. 4573  
November 2009

IZA

P.O. Box 7240  
53072 Bonn  
Germany

Phone: +49-228-3894-0  
Fax: +49-228-3894-180  
E-mail: [iza@iza.org](mailto:iza@iza.org)

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

## ABSTRACT

### Some Evidence on the Nature of Urbanization Economies<sup>\*</sup>

Urbanization economies – the effects on productivity and utility created endogenously by larger cities – are a fundamental component of both the economic geography of modern societies and the perpetuation of innovation and economic growth at a national level. Cities account for vast majorities of population – and even larger proportions of production and innovation – in all advanced economies. The nature of these endogenous effects of city size is thus of considerable importance. Krupka (2008) presents a general model in which exogenous variation in local productivity (“natural advantage”) and development constraints generate covariation in local incomes, housing prices and population. In that model, the strength of the correlation amongst these variables depends on the nature of the dominant urbanization economy (or diseconomy). This paper looks at the data over the last several decades and finds that the data is consistent with city size increasing consumer/resident happiness and/or *reducing* productivity of employers.

#### NON-TECHNICAL SUMMARY

Why do cities exist? What does the concentration of people and economic activity in cities add, or subtract, from production and well-being? This paper looks at how wages, housing prices and population relate to one another to gain insight into the effects of urbanization on people and business. I find that wages and housing prices are very tightly related, while both of these are less closely – but positively – related to city population. According to the theory described in this paper, this suggests that the primary effect of large population concentrations is to increase the happiness of residents, not increase the profits of businesses. These results are interesting because the ways in which cities increase happiness have been less closely examined in recent years.

JEL Classification: D5, J31, R12, R13, R23, R31

Keywords: agglomeration, urbanization economies, congestion, regional equilibrium, natural advantage, economic geography

Corresponding author:

Douglas J. Krupka  
Institute for Research on Labor, Employment and the Economy (IRLEE)  
University of Michigan  
506 E. Liberty St.  
Ann Arbor, MI 48104-2210  
USA  
E-mail: [dkrupka@umich.edu](mailto:dkrupka@umich.edu)

---

<sup>\*</sup> I would like to thank Lex Borghans and Armin Falk for helpful comments without implicating them in any errors or oversights I may have made.

## **I. Introduction**

The concentration of population and economic activity in cities is a prominent characteristic of all developed economies. Cities have always been centers of trade, growth and innovation. Because of their importance, the function of cities is of critical importance. Do cities arise due to pre-existing advantages to of certain locations? Do cities hamper production through the congestion of agents in space? Or, do cities generate advantages to local agents by bringing people together in close proximity?

These issues have attracted the interest of an army of economists, geographers and regional scientists over the years. Recent theories from economists (see Ottavian and Thisse, 2004, and Duranton and Puga, 2004, for recent reviews) have focused on ways that concentrations of people and commerce increase the productivity of firms or the workers they employ. In these models, cities can arise endogenously on a “featureless plain.” An older literature Tolley and Crihfield (1987) focused more on the negative effects of large cities: pollution, crime, traffic congestion, etc.

This paper uses the model I developed in an earlier paper (Krupka, 2008) to interpret the data on the covariance of metropolitan incomes, housing values and populations. That paper imbeds a reduced-form agglomeration economy in a neo-classical, semi-open city model to derive the general equilibrium effects of productive amenities (“natural advantage”) and housing supply constraints on equilibrium wages, housing prices and population of the city. In the presence of variation in both productive amenities *and* housing supply constraints, Krupka (2008) shows that wages should be more closely related to housing prices than population if cities mainly promote consumer satisfaction or reduce firm profitability, while wages should be more closely related to

population than housing prices if city size primarily increases profitability of firms. Pointing to a cross section of metropolitan average wages and home values for 2006, I suggested that the data appears to support the first case. Wages are more tightly correlated with housing prices than population, suggesting that – at least at the margin – concentration of economic activity tends to either decrease productivity of firms or increase the utility of residents.

These results are interesting because the lion's share of recent theoretical attention has been focused on the ways in which large cities promote production. Much less serious attention has been devoted to the ways in which cities might hamper productivity or generate higher quality of life for residents. While interesting, the use of simple metropolitan averages is undesirable since it fails to account for possible differences in the quality of workers and residences in various locations. In this paper, I not only control for observed quality, I make use of the Integrated Public Use Microsample data all the way back to 1940 to test the stability of this pattern. I find that the pattern found in my earlier paper using metropolitan averages carries through when controlling for observed characteristics, and that the pattern is persistent as far back as the data goes. The data also suggest that the relative effects of city size on productivity may be becoming less important with time.

The rest of the paper is organized as follows. Section II gives a brief review of the theoretical results in Krupka (2008). Section III describes the data and the empirical models. Section IV presents the results and section V concludes.

## II. Model: review of Krupka (2008)

In Krupka (2008), I present a “semi-open” city model, where businesses and residents have preferences (in the form of indirect utility and profit functions) over wages, rents and “natural advantage” amenities. These natural advantage amenities increase firm profits but do not affect residents’ utility. Wages increase utility and decrease profits, while rents decrease both utility and rents. The model is similar to that of Rappaport (2008a, b) in that it takes an individual city and “the rest of the economy” as its focus. Because the city is small relative to the economy, utility levels in the rest of the economy can be taken as exogenous. People and businesses locate in the city if their utility or profits exceed an individual-specific reservation level:

$$1) \quad \pi(r, w; A) \geq \varphi_j \sim F(.)$$

$$2) \quad v(r, w; A) \geq \gamma_j \sim G(.)$$

These reservation profit and utility levels are distributed (according to the CDFs  $F$  and  $G$ , respectively) because firms and people have different idiosyncratic attachments to and preferences for the city in question, and different outside options elsewhere in the national economy. In a pure “open city” model, these idiosyncratic factors would be zero for all actors. In a “closed city” model, they would be large enough to prevent migration of firms or people no matter what the local wages or rents were. The model falls somewhere between these extremes and can thus be considered to represent a semi-open city, which corresponds better to real cities than either pure case.

Equations 1 and 2 represent the locational equilibrium condition: no one must be willing to move in equilibrium. A second equilibrium condition is that the local labor market clears. In a simplified labor market where everyone works and each firm hires

exactly one worker,<sup>1</sup> this corresponds to the condition that the population of residents and of firms (or jobs) is equal:

$$3) \quad G(v(r, w; A)) - F(\pi(r, w; A)) = 0$$

Finally, each resident must have enough built space to live and work. In other words, the housing market must clear. The demand for housing is defined as the per capita demand for housing ( $D$ , which depends on wages and rents) times the residential population of the area ( $\Omega$ , which depends directly on utility, and thus indirectly on rents, wages and amenities). The supply of housing ( $S$ ) depends positively on rents and negatively on some supply constraint,  $C$ . Equilibrium rents are thus determined by the housing market clearance condition:

$$4) \quad D(r, w)\Omega(r, w; A) = S(r; C)$$

Equations 1-4 are sufficient to define equilibrium rent, wages and populations as functions of natural advantage amenities,  $A$ , and supply constraints,  $C$ . Intuitively, the labor market equilibrium condition specifies wages as a function of rents and amenities,  $w(r; A)$ . Given this function, it is straightforward to express population as a function of rents and amenities,  $\Omega(r; A)$ . Plugging these two expressions into equation four, we have an implicit function defining rent as a function of the two exogenous variables,  $r(A, C)$ , which can then be plugged back into the equilibrium wage and population functions to derive general equilibrium effects of the exogenous variables on the three easily observable variables: rents, wages and population.

Krupka (2008) then proceeds to include reduced-form urbanization effects into this model by allowing the utility of residents or the profits of firms to depend on the size

---

<sup>1</sup> The theoretical results are robust to more realistic assumptions about the labor market.

of the city in which they locate. It is possible to characterize and compare the effects of changes in natural advantage amenities and supply constraints on the endogenous variables under different assumptions about the effect of urban size on local firms and residents. The general equilibrium effects of the exogenous variables on the observable variables under different assumptions can then be compared to the actual covariation of the three observable, endogenous variables for insight into which kinds of urbanization forces are strongest at the margin in modern cities.

Three urbanization effects are considered in Krupka (2008). Productive congestion posits, along with Tolley and Crihfield (1987) that larger cities reduce productivity and profits of firms located there. On the other hand, larger cities might improve profit or productivity through a variety of channels. This type of urbanization effect is called “productive agglomeration” in Krupka (2008). This type of urbanization effect has dominated the recent agglomeration literature.<sup>2</sup> Finally, cities might make residents happier through the provision of better public goods or better variety of services. This type of urbanization effect is called consumer agglomeration. In my earlier paper, I find that none of these urbanization effects are enough to generate the clear positive correlation among rents, wages and populations we observe in the data. Allowing natural advantage amenities to vary among cities does induce the observed positive relationship amongst the three variables, although the effect is ambiguous for productive agglomeration.

If, however, city sites differ in terms of both natural advantage *and* supply constraints (which seems reasonable), there are some differences across the different

---

<sup>2</sup> The New Economic Geography literature posits two kinds of urbanization effects: a productive agglomeration effect of firms being closer to their output markets, and a consumer agglomeration effect of residents enjoying the agglomeration’s higher variety.



urbanization economies. The main difference between the three urbanization effects is that supply constraints increase wages (and rents) in the consumer agglomeration and production congestion models, but *decrease* wages (and population) in the production agglomeration model. While the variance in natural advantage induces a correlation between all three observable variables, the independent variation in supply constraints will tend to reinforce the correlation between wages and rents in the consumer agglomeration and production congestion models while it tends to reduce the correlation between wages and population. The results for productive agglomeration<sup>3</sup> are the opposite: the independent variation in housing supply tends to increase the correlation between wages and population while decreasing the correlation between wages and rents.

Thus, the models suggest that we should observe different patterns in the relative strength of the correlation amongst the three observable variables, depending on the kind of urbanization economy dominating at the margin in modern cities. In Krupka (2008), I show that the relationship between metropolitan average log incomes and log housing values is tighter than the relationship between average log incomes and log population, and interpret this as suggestive that – at the margin – the dominant form of urbanization economy tends to *decrease* profits and/or increase the *utility* of residents. These results are interesting because these are the forms of urbanization economy to attract the least attention in recent years, but they can only be taken so far since they do not control for housing or worker heterogeneity and are applied to only one year. The following section remedies this situation in both regards.

---

<sup>3</sup> A similar result would hold for consumer congestion.

### III. Empirical strategy and data.

The basic empirical test implied by Krupka (2008) is that the correlation between average rents and wages, conditional on observed characteristics of housing and workers ( $X$ ) will be higher than the correlation between population and average wages (conditional on worker characteristics) if consumer agglomeration and/or productive congestion dominate cities at the margins while the opposite will be true if productive agglomeration dominates around equilibrium. In symbols, Krupka (2008) suggest that inequality five will hold if consumer agglomeration or productive congestion dominate at the margin:

$$5) \quad \rho_{\bar{r}\bar{w}}|_X - \rho_{\bar{w}\Omega}|_X > 0$$

The limited empirical analysis in Krupka (2008) did not use conditional correlations. The strategy here is to compare metropolitan average log housing price and log income *conditional on individual characteristics* by computing the correlations between wage and housing price fixed effects from regressions which control for available observable characteristics.

The data comes from the Integrated Public Use Microdata Series (IPUMS), compiled by Ruggles et al. (2009). I use data from 2007, 2000, 1990, 1980, 1970 and 1940.<sup>4</sup> The available variables vary across the years. Most of the housing value regressions include controls for the number of rooms up to nine (*rooms*), an indicator variable for dwellings with more than eight rooms (*rooms9*), the number of bedrooms up to five (*bdrooms*), a dummy variable for dwellings with more than four bedrooms (*bdroom5*), indicator variables for lots over ten acres (*acre10*) complete kitchen (*ikitchen*), plumbing (*iplumb*), phone (*iphone*), publicly supplied water (*water*), septic

---

<sup>4</sup> The 1960 IPUMS data does not include information on metropolitan area of residence, while the 1950 IPUMS data does not include information on income or housing prices. The 1940 data includes some limited data on personal characteristics, but only very limited information on the dwelling.

tanks (*rawsewage*) and type of heating (*gasheat* and *elecheat*). There are also a set of variables for the vintage or age of the structure (*blt\**) and the type of unit (*un\**). Finally I include a set of controls for the utility costs of the dwelling (*lnutils*) and the commute time of the household head (*trantime*) and the average travel time of workers in the household (*tranavg*).

In the wage equations, I run standard Mincerian regressions controlling for experience and its square (*exper* and *exper2*) a set of educational attainment dummies (*edu\**), indicators for *male*, *white*, *black* and *hispanic*. Also included are indicator variables for whether the respondent is a citizen (*icitizen*), veteran (*veteran*) and has no disabilities preventing work (*able*). English proficiency is measured with two variables for being an English-only speaker (*english*) and speaks English “very well” (*bilingual*). The omitted categories are those who do not speak English, or can speak English, but not “very well.” Finally, I include controls for marital status (*married*, which does not include separated individuals) and either the presence of children in the household (*ikids*) or the number of children in the household (*kidsno*).

There are various options for both the housing and wage side of the analysis. For instance, in the housing equations the dependent variable could be either log of contract rent (*lnrent*), the log of the self reported housing value (*lnvalue*) or the log of the sum of these (*lnhousing*, so long as a control variable for whether the dwelling is rented or owned is included). For the wage equation, I use the log of wage and salary income (*lnwageinc*) as the dependent variable, but can choose between including the *ikids* or *kidsno* variable. There is also the choice between conditioning the sample on being a fulltime worker or including all workers *while controlling for the number of hours and*

*weeks worked in the previous year.* None of the conclusions drawn below are in the least sensitive to any of these decisions. As my preferred set of results I will present housing regressions based on both renters and homeowners (with a control for ownership status) and the wage equations which condition on fulltime work (over 49 weeks worked and at least 35 hours “usually” worked, or worked in the previous year) and use the children indicator. This choice is made because it is possible to run these regressions in every year of data I have. From year to year, some of these variables are unavailable, while others not discussed here become available. The available housing variables in 1940, for instance, are extremely sparse. In that year I use the number of families and subfamilies in the structure and family income (of the household head) as proxies for quality of the structure. The variables used in each year, along with the variable name and a brief description of the variable appear in Table 1.

After collecting the metropolitan fixed effects from the regressions discussed above, I correlate them with each other and with the log of the metropolitan population. There are also several choices for calculating this value. For recent years, census estimates for metropolitan population are electronically available. One problem with these estimates is that the metropolitan areas as defined by the census do not always match with those used to generate the fixed effects (which use the less detailed version of metropolitan area in the IPUMS data). An easy way around this is to take population estimates directly from the data used to estimate the fixed effects by counting up the person weights by metropolitan area in the IPUMS sample. These population estimates are correlated with the census estimates (when available) at over 0.95. My preferred population measure is that obtained by adding the person weights from the 1% PUMS in

each year because this measure is readily available whenever the analysis is possible. The choice of population measure does not affect any of the qualitative conclusions below.

#### **IV. Results**

Tables 2 and 3 present the results for the substantive variables for the housing and income equations for all years. Because the sample sizes are so large (at least several dozen thousand in every case), I do not report the robust standard errors or test statistics. Most variables return significant coefficients. In table 2, the sample includes both renters and owners, and is conditioned to housing units in metropolitan areas. Other sets of results using only renters or only owners return similar results. In table 2 we see the increase in the coefficient on *owner*, which shows the increase in the value-rent ratio over time, especially over the last few years before the recent troubles. Variables for rooms, bedrooms, kitchen and plumbing have expected signs (except for *rawsewage*). Also interesting is the substantial decrease in the otherwise steady coefficient of *iphone* in 2007, which might be interpreted as signaling the entrance of wireless and cable phone technology onto the stage.

Dummies for the vintage of the home show that newer homes (the excluded category) are worth more and that single family, detached homes (the excluded category) are worth more than mobile homes and attached homes in smaller structures, but that units in larger structures tend to be worth more than traditional suburban housing. This surprising result is probably due to selection; larger apartment buildings tend to be built only at very expensive locations. Finally, the coefficient on *tranavg* suggests that the

canonical urban economic model describes the data well. Homes located further from workplaces are worth less, all else equal. However, the often positive (but smaller) coefficient on the household head's commute time (*trantime*) shows that household heads who commute further to work tend to live in more expensive houses.

In 1940, when most of these housing quality controls are not available, we see that farms are worth less than non-farms (in this metropolitan-only sample). Housing units with more families but with fewer sub-families are worth more and that households headed by people with higher family incomes live in more expensive houses.

Table 3 presents the results from the models predicting the wage and salary income for household heads or their spouses who work full time (35 or more hours and at least 50 weeks in the previous year), are not self-employed and live in metropolitan areas. Results that include all metropolitan wage and salary workers and control for the amount of work done (instead of conditioning on full time work) return similar results. These results are for the most part unremarkable. Experience is rewarded with higher income at a decreasing rate. Education increases income (although there is some variance in whether associates and graduate degrees increase income relative to the next lowest level of education). Married, white males earn more, as do citizens, native English speakers and bilingual individuals (relative to people who have not yet mastered English). Disabled individuals make less and parents make more. Recently, veterans earned less income, although the opposite was the case as recently as 1990. At least in 1970, migrants made less, but children of migrants made more.

Table 4 presents the main results of the paper. In the top panel, it presents the relevant unconditional (or raw) correlations between metropolitan average "rents" and

“wages,” the raw correlation between metropolitan average wages and population, the difference between these two correlations, and the test statistic for that difference being different than zero. The bottom panel reports the same information for the conditional correlations: instead of metropolitan averages, these are correlations of the metropolitan fixed effects from the models reported in tables 2 and 3. Because there are a number of different possible specifications and measures, there is actually a distribution of such correlations. To give a sense of the sensitivity of the results to the specifications chosen, table 4 also reports the average across all possible pair-wise comparisons below the “preferred” correlations in each year (which correspond to the measures and models discussed above). With one exception (discussed below), none of the conclusions here are at all sensitive to the choice of model.

The top panel of Table 4 reproduces the analysis of Krupka (2008) for more years. There, we see that result reproduced: the correlation between metropolitan average wage and average housing price is higher than the correlation between average wage and population for every year considered. This is consistent with the dominant effect of increased city size being to increase residents’ happiness (consumer agglomeration). The result is also consistent with cities having grown to such sizes that they actually reduce firm profits (productive congestion). Looking at how the correlations change over the years, it appears that, if anything, this conclusion was stronger in the earlier time periods than more recently: city size effects were even more relatively consumer-oriented in the middle of the Twentieth Century than at the end, although the difference between the relative correlations is not too great (a 0.35 difference in correlations in 1940 versus 0.30 in 2007).

The bottom panel of Table 4 uses conditional correlations instead of raw correlations. These correlations between city fixed effects attempt to compare the wages and housing expenses of similar individuals across cities of different populations. The main result holds over from the unconditional analysis: the correlation between wages and rents is stronger than that between wages and population. This is consistent with either consumer agglomeration or productive congestion, but inconsistent with the productive agglomeration which has attracted the lion's share of recent research interest.

Looking at how the relative strength of these correlations have changed over time reveals a trend that was hidden in the unconditional analysis. In the bottom panel of Table 4, the difference between the two correlations appears to have grown over the later half of the Twentieth Century. The difference between the wage-rent conditional correlation and the wage-population conditional correlation increased by over 0.1 between 1940 and 2007.<sup>5</sup> The interpretation from the model is that, at the margin, cities have been becoming more relatively centered on the increase of residential utility over the Twentieth Century. While in the 1940's, this tendency is not that pronounced, by the end of our time period, the marginal effects of city size are centered very heavily either in the direction of increasing worker's utility or decreasing firms productivity. This characterization of the results holds no matter which specification, variable definition or population measure is used. The results are quite consistent in this regard.

---

<sup>5</sup> One might worry that the small difference in conditional correlations in 1940 is being driven by the inclusion of an income variable in the housing equation. The inclusion of an income variable in the housing equation would reduce the correlation between metropolitan income and housing price fixed effects. Running the analysis again without the family income variable gives results for 1940 very similar to those for 1970. Thus, the increase in the relative consumer-orientation of city size effects over the century is smaller with these results, although the qualitative differences remain. The appropriate numbers probably lie somewhere between those with and without the family income controls. On the one hand, including income reduces the correlation of wage and housing fixed effects, on the other hand, income is the only reasonable proxy for housing quality available in the data set.



There is one set of aberrant results (not reported in Table 4) which bears some discussion. In 1940, there is no significant difference between the two conditional correlations when housing expenses are measured using the value of home-owner's homes instead of rents (or the combination of rents and values as in the models presented above). If this were the only estimate of the correlations, it would suggest that in 1940 city size did not increase productivity any less than utility, which is in stark contrast to the rest of the results.<sup>6</sup> However, the data point fits in quite well with the rest of the results in terms of the changes in the relative effects of city size. This result, along with all the other results summarized in Table 4, is consistent with a dynamic in which the utility-effects of city size have been growing while the productivity-enhancing effects of city size have been becoming less important (or have actually been reversed) with time. Such a conclusion squares well with intuition about how increasing wealth should change residence decisions (it should make urban workers value urban amenities more), and the move towards a service-based economy.

## **V. Conclusion**

This paper has provided some empirical evidence on how the size of a city affects the well being of its resident workers and firms. Using the model from Krupka (2008) as a lens, I compare the conditional correlation between local housing prices and local incomes with the conditional correlation between local population and local incomes.

The finding that the housing-income correlation is stronger is consistent across years and across specifications. The implication of this empirical fact is that city size tends not to

---

<sup>6</sup> The difference is actually negative, -0.013, with a test statistic of -0.15. In the models that exclude family income from the set of control variables the difference is larger at 0.115, but still insignificant with a test statistic of 1.46.

increase business productivity or profits. Instead, larger cities tend to either increase the wellbeing of their residents or actually *decrease* the profitability of resident businesses.

These results fit in well with some recent papers. Rappoport (2008a) looks at “consumption amenities” and finds variance in such amenities is sufficient to support the variance in city crowdedness observed in the data. Rappoport (2008b) comes to the opposite conclusion from the perspective of productivity differences. Looking at differences between less and more educated individuals, DalMazzo and Blasio (2009) find evidence for Italian cities consistent with larger cities providing utility benefits at least to more educated individuals. This paper adds to those results by looking at the implied covariation in wages, rents and population. The mounting evidence of the importance of consumption amenities in the agglomeration of population and economic activity in large cities suggests that they warrant further theoretical and empirical investigation.

#### References:

- DalMazzo, Alberto and Guido de Blasio (2009). “Amenities and skill-biased agglomeration effects: Some results on Italian cities.” Unpublished working paper.
- Duranton, Gilles and Diego Puga (2004). “Micro-foundations of Urban Agglomeration Economies.” chp. 48 in *Handbook of Regional and Urban Economics*, vol. 4, V. Henderson and J.F. Thisse, eds. Amsterdam: North-Holland: 2063-2117.
- Krupka, Douglas J. (2008). “On amenities, natural advantage and agglomeration.” IZA Discussion paper no. 3598 available at: <http://ftp.iza.org/dp3598.pdf>
- Ottaviana, Gianmarco I.P. and Jacques-Francois Thisse (2004). “Agglomeration and economic geography,” chp. 58 in *Handbook of Regional and Urban Economics*, vol. 4, V. Henderson and J.F. Thisse, eds. Amsterdam: North-Holland: 2563-2608.

Rappaport, Jordan (2008a). "Consumption amenities and city population density." *Regional Science and Urban Economics* **38**: 533-552.

Rappaport, Jordan (2008b). "A productivity model of city crowdedness." *Journal of Urban Economics* **63**: 715-722.

Ruggles, Steven, Matthew Sobek, Trent Alexander, Catherine A. Fitch, Ronald Goeken, Patricia Kelly Hall, Miriam King, and Chad Ronnander. *Integrated Public Use Microdata Series: Version 4.0* [Machine-readable database]. Minneapolis, MN: Minnesota Population Center [producer and distributor], 2009. Available at: <http://usa.ipums.org/usa/>

Tolley, George S. and John Crihfield (1987). "City size and place as policy issues." chp. 33 in *Handbook of Regional and Urban Economics*, vol. 2, Edited by E.S. Mills Amsterdam: North-Holland: 1285-1311.

**Table 1: Variable Definition and availability**

Variable name	Short Description	Inclusion Year					
		2007	2000	1990	1980	1970	1940
Inrent	log of contract rent	Yes	Yes	Yes	Yes	Yes	Yes
Invalue	log of home value	Yes	Yes	Yes	Yes	Yes	Yes
Inhousing	sum of Inrent and Invalue	Yes	Yes	Yes	Yes	Yes	Yes
owner	Home Onwership Dummz	Yes	Yes	Yes	Yes	Yes	Yes
rooms	Number of rooms	Yes	Yes	Yes	Yes	Yes	
rooms9	Indicator for 9 or more rooms	Yes	Yes	Yes	Yes	Yes	
bdrooms	Number of Bedrooms	Yes	Yes	Yes	Yes		
bdroom5	Indicator for 5 or more bedrooms	Yes	Yes	Yes	Yes		
ikitchen	Kitchen Dummy	Yes	Yes	Yes	Yes	Yes	
iplumb	Complete Plumbing Dummy	Yes	Yes	Yes	Yes	Yes	
ihotwat	Hot Water dummy					Yes	
ibathe	shower or tub dummy					Yes	
icrap	Toilet dummy					Yes	
bath10	One full bath dummy					Yes	
bath15	1.5 bath dummy					Yes	
bath2p	Two or more bath dummy					Yes	
ibase	Basement dummy					Yes	
rawsewage	Septic Tank dummy			Yes	Yes	Yes	
water	Piped in water dummy			Yes	Yes	Yes	
iphone	Available phone dummy	Yes	Yes	Yes	Yes	Yes	
centair	Central Air Conditioning dummy					Yes	
gasheat	Natural gas heating dummy	Yes	Yes	Yes	Yes		
elecheat	Elevtric heating dummy	Yes	Yes	Yes	Yes		
Inutils	log of utility payments	Yes	Yes	Yes	Yes	Yes	
blt__ <sup>a</sup>	Housing age dummies	Yes	Yes	Yes	Yes	Yes	
unmob	Mobile home dummy	Yes	Yes	Yes	Yes	Yes	
un1det	Single family, detached	Yes	Yes	Yes	Yes	Yes	
un1att	Single family, atached	Yes	Yes	Yes	Yes	Yes	
un__ <sup>b</sup>	Units in structure dummies	Yes	Yes	Yes	Yes	Yes	
acre10	On lot of more then 10 acres	Yes	Yes	Yes	Yes	Yes	
trantime	HH-head's commute time	Yes	Yes	Yes	Yes		
tranavg <sup>c</sup>	Average commute at unit	Yes	Yes	Yes	Yes		
NoFam	Number of families in Household						Yes
NoSubFam	Number of subfamilies in Household						Yes
FamInc	Family income						Yes

**Table 1 (cont.)**

Variable name	Short Description	2007	2000	1990	1980	1970	1940
lnwageinc	log of wage and salary income	Yes	Yes	Yes	Yes	Yes	Yes
lnweeks	log of weeks worked, previous year	Yes	Yes	Yes	Yes		Yes
lnhours	log of hours usually worked	Yes	Yes	Yes	Yes		Yes <sup>d</sup>
exper	Experience	Yes	Yes	Yes	Yes	Yes	Yes
exper2	Experience-squared	Yes	Yes	Yes	Yes	Yes	Yes
eduNoHS	Education less than highschool	Yes	Yes	Yes	Yes	Yes	Yes
eduHSDrop	Highschool drop out	Yes	Yes	Yes	Yes	Yes	Yes
eduHS	Highschool graduate	Yes	Yes	Yes	Yes	Yes	Yes
eduScol	Some college, no degree	Yes	Yes	Yes	Yes <sup>e</sup>	Yes <sup>e</sup>	Yes <sup>e</sup>
eduAss	Associates degree	Yes	Yes	Yes			
eduBach	Bachelors degree	Yes	Yes	Yes	Yes <sup>f</sup>	Yes <sup>f</sup>	Yes <sup>f</sup>
eduGrad	Graduate Degree	Yes	Yes	Yes			
male	Male Dummy	Yes	Yes	Yes	Yes	Yes	Yes
married	Married (not seperated)	Yes	Yes	Yes	Yes	Yes	Yes
white	White dummy	Yes	Yes	Yes	Yes	Yes	Yes
black	Black dummy			Yes	Yes	Yes	Yes
hispanic	Hispanic dummy	Yes	Yes	Yes	Yes		Yes
NatEnglish	Native English speaker					Yes	
icitizen	Citizen dummy	Yes	Yes	Yes	Yes		
english	Speaks only english	Yes	Yes	Yes	Yes		
bilingual	Speaks english "very well"	Yes	Yes	Yes	Yes		
bornhere	born in USA					Yes	Yes
AmeriMom	Mother born in USA					Yes	
AmeriDad	Father born in USA					Yes	
able	No disability prevents work	Yes	Yes	Yes	Yes		
veteran	Veteran dummy	Yes	Yes	Yes	Yes	Yes	Yes
kidsno	Number of children in household	Yes	Yes	Yes	Yes	Yes	Yes
ikids	Children present in household dummy	Yes	Yes	Yes	Yes	Yes	Yes

- a Categories include new, 2-4, 5-9,10-19 and further ten year increments up to 60 plus., except in 2007, where the variable is based on decade of construction, effectively adding two years to all bins from 4 up.
- b Categories include 2, 2-4, 5-9, 10-19, 20-39 and 50 plus.
- c Includes HH Head and spouse , if present and working. If there is no spouse or if spouse does not work, this is the same as trantime.
- d Hours worked in previous week used instead of hours "usually worked last year"
- e Coding varies in earlier years: Some college includes all education between 12 and 16.
- f Coding varies in earlier years: All college degree holders, including graduate degrees are coded here.

**Table 2: Housing coefficients across years**

Variable	2007 coefficients	2000 coefficients	1990 coefficients	1980 coefficients	1970 coefficients	1940 coefficients
owner	5.4616	5.1843	5.1340	5.2938	4.9591	4.8067
rooms	0.0969	0.0927	0.1184	0.1270	0.0792	
rooms9	0.1491	0.2057	0.1462	0.1319	0.0655	
bdrooms	0.0632	0.0362	0.0249	0.0048		
bdroom5	0.0176	0.0906	0.0410	0.0271		
ikitchen	0.0785	0.0940	0.0396	0.0856	0.0291	
iplumb	0.0065	0.0680	0.2168	0.2224	-0.3379	
ihotwat					0.4111	
ibathe					0.2483	
icrap					0.1328	
bath10					0.2671	
bath15					0.4816	
bath2p					0.6143	
ibase					0.1122	
rawsewage			0.0647	-0.0457	0.0092	
water			-0.0529	0.0088	0.0101	
iphone	0.0275	0.1830	0.1878	0.1432	0.1347	
centair					0.2336	
gasheat	0.0077		-0.0204	-0.0105		
elecheat	0.0177		0.0463	0.0134		
lnutls	0.0305		0.0137	-0.0038	-0.0093	
blt2-4 <sup>a</sup>	-0.0520	-0.0637	-0.0423	-0.0521	-0.0412	
blt5-9	-0.1242	-0.1418	-0.1609	-0.1526	-0.1028	
blt10-19	-0.2353	-0.2474	-0.2461	-0.2332	-0.1668	
blt20-29	-0.3255	-0.3596	-0.3104	-0.3565	-0.2816	
blt30-39	-0.3590	-0.4076	-0.3640		-0.4497	
blt40-49	-0.3874	-0.4472	-0.4383	-0.5590		
blt50-59	-0.4234		-0.4427			
blt60plus	-0.3759	-0.4378				
unmob	-1.2902	-1.2130	-0.9664	-0.2188	-0.1220	
un1att	-0.1026	-0.1527	-0.1195	-0.1488	-0.1926	
un2	-0.0378	-0.0754	-0.0238	0.0088	-0.0308	
un3-4	-0.0452	-0.0385	-0.0201	0.0175	-0.0139	
un5-9	-0.0271	-0.0129	-0.0134	0.0618	0.0153	
un10-19	0.0117	0.0210	0.0097	0.0963	0.0630	
un20-29	0.0024	0.0177	0.0138	0.1079	0.0871	
un50plus	0.1197	0.1231	0.0536	0.1947	0.1660	
acre10	0.3369	0.3043	0.0762	0.1298	-0.0157	
trantime	-0.0004	0.0003	0.0010	0.0013		
tranavg	-0.0008	-0.0012	-0.0017	-0.0018		
ifarm						-0.2714
nfams						0.1770
nsubfam						-0.0111
lfaminc						0.3468

<sup>a</sup> In 2007, these variables are coded according to decade of construction, so that the numbers are not exactly the same.

**Table 3: Income coefficients across years**

Variable	2007 coefficients	2000 coefficients	1990 coefficients	1980 coefficients	1970 coefficients	1940 coefficients
exper	0.0330	0.0306	0.0317	0.0301	0.0250	0.0386
exper2	-0.0005	-0.0005	-0.0005	-0.0005	-0.0004	-0.0005
eduHSDrop	0.0242	0.0557	0.0750	0.0695	0.1151	0.1805
eduHS	0.1763	0.2048	0.2220	0.2162	0.2450	0.3427
eduScol	0.3405	0.2143	0.2140	0.3406	0.3797	0.4680
eduAss	0.4248	0.1308	0.1482			
eduBach	0.7135	0.4006	0.3651	0.5943	0.6577	0.6808
eduGrad	0.9586	0.2528	0.2624			
male	0.2967	0.3140	0.3681	0.4486	0.5058	0.5105
married	0.0792	0.0739	0.0556	0.0508	0.0498	0.0649
white	0.1448	0.1314	0.0787	0.0691	0.1160	0.5161
black			-0.0881	-0.1291	-0.1828	-0.1214
hispanic	-0.0836	-0.0819	-0.0813	-0.1235		-0.2046
NatEnglish					0.0555	
bornhere					0.0848	0.1255
AmeriMom					-0.0075	
AmeriDad					-0.0199	
icitizen	0.1170	0.0956	0.1255	0.0806		
english	0.2868	0.2092	0.1895	0.1586		
bilingual	0.2527	0.1845	0.1462	0.1295		
able	0.1319	0.0664	0.1475	0.1243		
veteran	-0.0112	-0.0113	0.0233	0.0490	0.0226	0.0514
ikids	0.0343	0.0239	0.0135	0.0198	0.0401	0.0252

**Table 4: Raw and conditional correlations, across years**

<b>Raw</b>		Year					
		2007	2000	1990	1980	1970	1940
$\rho_{r\bar{w}}$	Pref.	0.7566	0.8372	0.7289	0.5751	0.7411	0.7572
	Avg.	0.7576	0.8056	0.7251	0.5658	0.7277	0.7205
$\rho_{w\Omega}$	Pref.	0.4613	0.5674	0.4383	0.3152	0.4769	0.4033
	Avg.	0.4742	0.5711	0.4749	0.3324	0.4769	0.4093
$\rho_{r\bar{w}} - \rho_{w\Omega}$	Pref.	0.2953	0.2698	0.2906	0.2599	0.2642	0.3539
	Avg.	0.2834	0.2345	0.2502	0.2335	0.2508	0.3112
z	Pref.	5.80	3.87	5.29	3.80	3.40	4.50
	Avg.	5.66	3.25	4.62	3.39	3.18	4.00
<b>Conditional</b>							
$\rho_{r\bar{w}}$	Pref.	0.8038	0.8350	0.7539	0.6236	0.8038	0.7070
	Avg.	0.8137	0.8314	0.7647	0.6049	0.7814	0.6616
$\rho_{w\Omega}$	Pref.	0.5364	0.5458	0.5025	0.3856	0.5945	0.5488
	Avg.	0.5388	0.6281	0.5261	0.3886	0.5952	0.5432
$\rho_{r\bar{w}} - \rho_{w\Omega}$	Pref.	0.2674	0.2892	0.2514	0.2380	0.2093	0.1582
	Avg.	0.2749	0.2032	0.2386	0.2163	0.1862	0.1184
z	Pref.	6.05	4.04	4.98	3.75	3.33	2.12
	Avg.	6.40	3.11	4.85	3.36	2.87	1.60