An empirical index for labor market density*

Pieter A. Gautier[†]
Coen N. Teulings

October 16, 2002

Abstract

In this paper we derive a structural index for labor market density based on the Ellison and Glaeser (1997) index for industry concentration. The labor market density index serves as a proxy for the number of workers that is potentially available for jobs in a particular area. The index is based on observed home-work location patterns. It is particularly useful for testing theories where the scale of the market matters. We apply this index to a standard wage equation and find that it explains almost half of the cross- region-wage-variance.

Keywords: labor market density, wage equation
JEL codes: J210, J300, J600, J230

*Part of this research has been carried out at the Industrial relations office in Princeton. We thank two anonymous referees, Glenn Ellison, Alan Krueger and participants at the Tinbergen Institute and the ESEM 2001 in Venice for useful comments and David Jaeger for kindly providing us with his program to map county groups to (C)MSA's.

[†]Erasmus University Rotterdam and Tinbergen Institute; gautier@tinbergen.nl, teulings@few.eur.nl

1 Introduction

Search frictions play an important role in the labor market. Job seekers and vacancies do not meet instantaneously, their matching takes effort and time. The efficiency of this matching process depends on the characteristics of the labor market. An obvious factor that matters is the density of the market: the more job seekers and vacancies are available in a particular area, the easier it is for them to find an acceptable match. Several authors have developed empirical models along these lines, see e.g.: Diamond (1982), Burda and Profit (1996), Coles and Smith (1998), Wasmer and Zenou (1999), Wheeler (2002) and Glaeser and Maré (2001). Although there is a large literature that suggests that returns to scale in job search are constant, there is a good reason why the number of job seekers and vacancies might matter. A larger labor market allows workers and firms to be more choosy so it reduces mismatch. This effect is typically ignored in the empirical matching literature which is based on aggregated time series data.

A big obstacle in research in this area is that labor market density is difficult to measure. One likely candidate is simply the amount of workers and/or jobs per square mile. However, a number of serious drawbacks to this measure immediately come to mind. First, it ignores the role of infrastructure. What we are really interested in is not the set of applicants within a certain distance of the job, but in the amount of workers that is potentially available for a job in a certain region. The relevant labor market area should then be weighted by the number of highways and public transport facilities. Moreover, when particular regions are more attractive as residential area, people might be prepared to accept on average a longer commuting time.

These considerations suggest that we should look for an index based on revealed preferences. The index that we propose is based on observed home-work location patterns. The idea is that we take the location of the job of a worker as given and then analyze where that worker lives. To clarify this idea, assume for the sake of argument that the economy consists of a number of areas with an equal number of inhabitants, say n. Then, if we observe that all workers live in the same area as where they work, a given job can only be occupied by n workers. This is typical for a small scale labor market. Alternatively, when workers working in a particular location live in 10 different areas than $10 * n$ workers can potentially take jobs in that area and the scale of the labor market is large. More specifically, our index can be viewed of as a model based index of geographic labor market density similar to the dartboard index for industry concentration of Ellison and Glaeser (1997, EG from now onwards). The index can take any value between zero and one. When it is equal to one, the only workers who work in a particular area are the ones who live there. When it is equal to zero, the labor market is extremely dense and we observe workers from many different areas to be employed in this labor market. In other words, when many workers are available for a particular job in a particular area we call this area dense.

The plan of the paper is as follows. Section 2 derives the index from location decisions of utility maximizing agents. Section 3 describes how the index can be constructed from the 5% PUMS of the Census and how it can be linked to the (C)MSA areas of the CPS. Finally, section 4 gives an illustration of how the index can be applied in a wage equation and in a model for the cost of living index. We find that 45% of the regional variation is captured by our density index.

2 The index

This section presents the density index which is a special case of EG's index for industry concentration. Consider the decision problem for the k th worker with a job in area w who has to choose an area h_k to live in. We take the distribution of jobs across areas as given and define f_w as the fraction of jobs located in area w. Let the utility for area h be given by:

$$\log \pi_{kwh} = \log \bar{\pi}_{wh} + \varepsilon_{kwh} \tag{1}$$

where the ε_{kwh}'s reflect idiosyncratic factors (like the relative preference for clean air, safety, theater availability etcetera) which are assumed to be independent Weibull random variables which are also independent of $\{\bar{\pi}_{wh}\}$, and $\bar{\pi}_{wh}$ is a random location specific variable, which is chosen by nature at the start of the process. It reflects the attractiveness to live in a certain area (given that the agent's job is in w) for a typical agent. Conditional on the realization of the random variables: $\bar{\pi}_{w1}, ..., \bar{\pi}_{wH}$ and given our assumptions on ε_{kwh} we can write the probability that an agent chooses area h as:

$$Prob\{h_k = h|\bar{\pi}_{w1}, ..., \bar{\pi}_{wH}\} = \frac{\bar{\pi}_{wh}}{\sum_j \bar{\pi}_{wj}} \equiv p_{wh}$$

$$Prob\{h_k = h|\overline{\pi}_{w1}, ..., \overline{\pi}_{wH}\} = \frac{\overline{\pi}_{wh}}{\sum_j \overline{\pi}_{wj}} \equiv \left(p_{wh}|\overline{\pi}_{w1}, ..., \overline{\pi}_{wH}\right)$$

which is a conditional logit model, see McFadden (1973). Note that p_{wh} is a random variable because $\overline{\pi}_{wj}$ are random variables. We assume that the distribution of p_{wh} is such that:

$$\sum_w f_w p_{wh} = \mu_h \tag{2}$$

$$var(p_{wh}) = \gamma_w \mu_h (1 - \mu_h) \tag{3}$$

where $\mu_h \in [0, 1]$ and $\gamma_w \in [0, 1]$. Equation (2) implies $E(p_{wh}) = \mu_h$, where the expectation is taken over individuals k. Let x_h be the fraction of the total population that chooses to live in area h. Then:[1]

$$E(x_h) = \sum_w f_w p_{wh} = \mu_h$$

The parameter μ_h drives the overall distribution of workers' home location across areas h. Hence, $\sum_h \mu_h = 1$. An appropriate choice of μ_h reproduces the distribution of home locations that is actually observed in the data, i.e. it puts more workers in New York than in Kansas. The variance of p_{wh} measures how sensitive the agent's utility is to region specific factors. For jobs in non-dense areas, the variance is likely to be high because, given the location of the job, there are few areas that have a sufficiently high value of $\overline{\pi}_{wh}$ while the rest of the areas have $\overline{\pi}_{wh} = 0$. When $\gamma_w = 1$, the variance of p_{wh} reaches a maximum (since the maximum of the variance of a random variable with support between zero and one with mean $\mu_h \in [0, 1]$ is $\mu_h [1 - \mu_h]$). In that case, the variation in idiosyncratic characteristics ε_{kwh} is dominated by the variation in the location specific factors, $\log \overline{\pi}_{wh}$. When $\gamma_w = 0$, the location decision is totally dominated by the agent's idiosyncratic taste factors. Region specific factors do not matter. This is the case in a fully integrated, dense labor market. The agent's decision on where to live is independent of the location of the job and each living area h is chosen with probability x_h. The parameter γ_w can therefore be interpreted as a density index for region w. In other words, γ_w captures the importance of regional factors relative to idiosyncratic taste factors of the agents.

[1] Equation (2) is slightly more general than the condition $E(p_{wh}) = x_h$ applied by EG. The latter condition implies that x_h is non-stochastic. This seems to be inconsistent with their model, since x_h depends on the realizations of ε_{kwh} and must therefore be stochastic. However, this difference in presentation has no implications for the Proposition1 below.

Now we will define an unbiased estimator for γ_w. Let s_{wh} be the number of workers working in area w and living in area h as a share of total employment in area w. Then:

Proposition 1 *Consider the case where K workers distributed across work locations w with share f_w choose their home location according to equations (1), (2), and (3). A consistent estimator for γ_w is:*

$$\underset{K\to\infty}{plim} \frac{\sum_h (s_{wh} - x_h)^2}{(1 - \sum_h x_h^2)} \tag{4}$$

Proof.
We have:

$$\begin{aligned}
\underset{K\to\infty}{\text{plim}} \left(\sum\nolimits_h (s_{wh} - x_h)^2 | p_{wh}\right) &= \underset{K\to\infty}{\text{plim}} \sum\nolimits_h E\left[(s_{wh} - x_h)^2 | p_{wh}\right] \\
&= \underset{K\to\infty}{\text{plim}} \sum\nolimits_h \left[var(s_{wh} - x_h | p_{wh}) + E^2\left(s_{wh} - x_h | p_{wh}\right)\right] \\
&= \sum\nolimits_h \left(p_{wh} - \mu_h\right)^2
\end{aligned}$$

The second step uses $var(s_{wh} - x_h|p_{wh}) = E\left[(s_{wh} - x_h)^2|p_{wh}\right] - E^2(s_{wh} - x_h|p_{wh})$. The third step uses $\text{plim}_{K\to\infty} var(s_{wh} - x_h|p_{wh}) = 0$, $E\left(s_{wh}|p_{wh}\right) = p_{wh}$ and $E\left(x_h\right) = \mu_h$. Dropping the conditioning on p_{wh} and substitution of (2) and (3) yields:

$$\begin{aligned}
\underset{K\to\infty}{\text{plim}} \left(\sum\nolimits_h (s_{wh} - x_h)^2\right) &= E_{p_{wh}}\left[\sum\nolimits_h \left(p_{wh} - \mu_h\right)^2\right] = \sum\nolimits_h var\left(p_{wh}\right) \\
&= \sum\nolimits_h \gamma_w \mu_h (1 - \mu_h) = \gamma_w \left(1 - \sum\nolimits_h \mu_h^2\right)
\end{aligned}$$

where we use: $E\left(p_{wh}\right) = \mu_h$ in the second equality and $\sum_h \mu_h = 1$ in the final equality. Rearranging terms and using $E\left(x_h\right) = \mu_h$ gives (4). ■

To illustrate how this index is related to the scale of the labor market, consider a job in area w. Let there be N residential areas, each populated by a single worker and let n be the number of workers who is willing to work in area w and let all of them have an equal probability to get this job. Hence, n is a measure for the scale of the labor market. The probability for each of the workers to get this job is $1/n$ and the probability for the rest of the population, $N - n$, to get the job is equal to zero. In other words, $p_{wh} = 1/n$ with probability n/N and $p_{wh} = 0$ with probability: $(1 - n/N)$. Since the variance of a Bernoulli trial with success rate: n/N is $(1-n/N)n/N$, the variance of p_{wh} is: $V = (1/n)^2[(1-n/N)n/N] = 1/N[1/n - 1/N]$. According to equation (3), this is equal to: $\gamma \frac{1}{N}(1-\frac{1}{N})$. Solving for γ and taking $lim\ N \to \infty$ gives $\gamma \simeq \frac{1}{n}$. Hence, in this simple binomial example where workers either do or do not belong to a market for a particular job and where all workers in a market have an equal probability for that job, γ is equal to the inverse of the scale of the labor market.

The above analysis takes as a starting point the work area of the worker and then determines the choice of the optimal living area. We could also have proceeded the other way around, by analyzing the choice of the optimal work area conditional on the living area. The actual conditioning on work area in our calculations is based on the notion that a large fraction of city centers consists of offices. Then, conditioning on living area would underestimate the density of the city centers. Most people living in Manhattan are likely to work in Manhattan, incorrectly suggesting that Manhattan is a low density area. However, most people working in Manhattan live in other regions. Hence, by conditioning on work areas we avoid the problem of the mismeasurement of γ_w in city centers.

Under the assumptions made, this index is independent of its level of aggregation.[2] Whether one measures location at for example the state level or the county level should not affect the calculated value of γ_w for a state. However, this requires that the values of $\{\bar{\pi}_{wh}\}$ are drawn independently of the aggregation scheme of subregions into regions. Obviously, this assumption is violated in reality. In practice, any aggregation merges adjacent sub-regions into a new region. The values of $\{\bar{\pi}_{wh}\}$ for sub-regions within a region will typically be correlated. The example below makes this clear.

Consider 4 regions of equal size ($x_h = 1/4$), each consisting of 4 agents and 4 jobs. The first case represents the case where all regions form a fully integrated market; $s_{wh} = 1/4$ for all h, w and $\gamma_w = 0$. In the second case, $s_{wh} = 1/2$ for all h, w, and $\gamma_w = 1/3$. This is typical for the situation where 1 and 2 as well as 3 and 4 are twin cities. In the third case there are 4 fully separated markets: $s_{wh} = 1$ for all h, w and $\gamma_w = 1$.

[2] There are two ways to think about aggregation in this context: (1) reducing the number of areas for which we calculate γ_w and (2) taking weighted averages of γ_w over neighboring areas. The discussion in this section is about (1) while in section 4 we consider (2).

(1) $\gamma_w = 0$				
$w \setminus h$	1	2	3	4
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1
4	1	1	1	1

(2) $\gamma_w = 1/3$				
$w \setminus h$	1	2	3	4
1	2	2	0	0
2	2	2	0	0
3	0	0	2	2
4	0	0	2	2

(3) $\gamma_w = 1$				
$w \setminus h$	1	2	3	4
$w1$	4	0	0	0
$w2$	0	4	0	0
$w3$	0	0	4	0
$w4$	0	0	0	4

When we combine regions 1 and 2 and regions 3 and 4 into two new regions, we get:

(1) $\widehat{\gamma}_w = 0$		
$w \setminus h$	$h1,2$	$h3,4$
$1,2$	4	4
$3,4$	4	4

(2) $\widehat{\gamma}_w = 1$		
$w \setminus h$	$1,2$	$3,4$
$1,2$	8	0
$3,4$	0	8

(3) $\widehat{\gamma}_w = 1$		
$w \setminus h$	$1,2$	$3,4$
$1,2$	8	0
$3,4$	0	8

When $\gamma_w = 1/3$, combining regions 1 and 2 increases γ_w .[3] The extreme cases are invariant to the aggregation of regions. For the other cases, aggregation tends to overestimate γ_w. In the next section, we present estimates of γ_w for the US and we will test whether aggregation affects the results.

3 Data

3.1 Constructing the index from Census data

The US Census data are well suited for the construction of our index because they contain detailed information on both the area of residence and the work area at low levels of aggregation. We use the 5% public use micro samples (PUMS) of the 1990 Census. The most disaggregate geographic unit in the Census is the Public Use Micro data Area (PUMA). A typical PUMA is populated by at least 100,000 persons and is identified by a five-digit number which is unique within states. In dense areas, PUMA's define a subset of a single county while in the rural states, PUMA's typically consist of several different counties. To construct our density index we also need information on the area where the worker works (PUMAW). This is however defined at the 2 digit level (unique by state) which will be the level of our analysis. With the method of the previous section we were able to construct a γ_w for each of the 1138 2-digit PUMA's.

[3] Note that if we would have combined regions 1,4 and 2,3 or 1,3 and 2,4 this would reduce γ_w but in practice this is not relevant because aggregation schemes tend to combine adjacent and integrated sub-regions.

In calculating γ_w, we only included the full time employed workers and excluded Alaska and Hawaii. Since in general, each area is very small compared to the whole country, the denominator of (4) is close to one (i.e. using Census data, we found for the US: $\sum_w x_h^2 = 0.0024$) and γ_w is therefore almost entirely determined by $\Sigma_h (s_{wh} - x_h)^2$. To get an idea of the range of possible values γ_w can obtain, we found γ_w to be equal to 0.07 in Northern New Jersey while for some areas in Arizona, Maine, Missouri, Montana, Kansas and Wyoming we found values of γ_w as high as 0.95. The size distribution is plotted in Figure 1. Both the mean and the standard deviation of the relative PUMA sizes are 0.001. It suggests that we do not have to worry about aggregation bias. This is confirmed by a simple OLS regression of log γ_w on the log relative size of the area which shows that there exists an insignificant positive relation between γ_w and area size (the elasticity $= 0.02$, $t = 0.56$).

3.2 Using additional information from the CPS

For many economic applications, the CPS contains crucial individual information which is not present in the Census. That is why we aggregated up our index to the (C)MSA*state level. This is not a trivial operation because there is no one to one match between the PUMA's of the Census and the (C)MSA (central metropolitan area) and state classification of the CPS. We therefore use the following strategy. First, we match the PUMAW's to (C)MSA 's, using the method of Jaeger et al. (1997). We aggregate by taking weighted (by relative area size) averages of the relevant γ_w's. In most states there are however areas which do not belong to a (C)MSA. Those are typically rural areas. For those areas we also calculated weighted average γ_w $'s$ per state.[4] Finally there are some small (C)MSA's that consist of only 1 PUMA. When those areas are isolated (i.e. Tuscon, Phoenix) or close to the Mexican border (i.e. El Paso), this overestimates γ_w for the reasons we discussed in section 2. We therefore treat (C)MSA's consisting of only one PUMA as the within state areas that do not belong to a (C)MSA. This leaves us with in total 164 γ_w's. To illustrate the aggregation procedure, consider the following example for Indianapolis, IN. The Indianapolis CMSA, consists of four PUMA's, each with a unique γ_{Census}. In the CPS, Indianapolis is treated as a single geographical unit. We take weighted (by x_w)

[4]For the definitions of (C)M(S)A's we refer to the apendix. Our density measures and relevant weights per PUMAW of the 1990 census and per (C)MSA/MA of the CPS, and SAS formats for (C)MSA's and states can be found at: http: // www.tinbergen.nl / ~gautier / lmdensity.html.

averages of γ_{Census} to get a unique γ_{CPS} for Indianapolis.

Figure 2 plots density distributions for both the 1138 Census PUMA's and the 164 CPS areas. The mean for the Census γ_{Census} is 0.597 and the standard deviation is 0.235 while for γ_{CPS} those values are respectively 0.574 and 0.185. From this, we conclude that we do not loose much variation in our index by aggregating up to the (C)MSA*state level, suggesting that the CPS regions are quite homogeneous with respect to their γ. The overall shape of the distribution is quite similar, it is bimodal with one hump at $\gamma = 0.80$, while the other hump is at $\gamma = 0.25$ for the Census and at $\gamma = 0.40$ for the CPS.[5]

Figure 2 about here

We expect γ_w to be related to population density (measured in persons per square mile). Figure 3 is illustrative in this respect. Figure 3 shows a map of all the counties in the U.S., where the darker areas are more densely populated. In this Figure we inserted some values of γ_w, based on the Census public use micro areas. We clearly see that densely populated areas have smaller $\gamma_w's$. The correlation between γ_{CPS} and people per square mile is -0.43.

Figure 3 about here

4 Application: Estimation of a wage equation

In this section we look at the effect of our labor market density index on wages. This application merely serves as an illustration. We do not have a narrow structural interpretation of our estimation results. In the literature, several reasons for the existence of cross regional wage differentiation have been put forward like regional differences in the efficiency of the matching process as in Teulings and Gautier (2001) and Wheeler (2002), knowledge spillovers as in Lucas (1988) and Glaeser and Maré (2001) or compensating differentials as in Roback (1982). For our purposes it is enough that wages are correlated with labor market density. We are interested in the fraction of the cross regional variance in wages that can be attributed to our labor market density index. If density matters, it should pick up a substantial part of the cross-regional variation in wages. We use both the March 1991 supplements of the CPS and the 1% PUMS of the 1990 Census for our wage equation. Most of our attention goes to the CPS results because it allows for a

[5] See the appendix with a full list of γ.

more accurate calculation of earnings and hours worked. In the PUMS, working time was measured as an interval variable which makes the hourly wage rate less accurate.[6]

Directly estimating the effect of γ on log wages with OLS gives an unbiased estimate of λ but it produces downwardly biased standard errors in the presence of within-region-correlation of the disturbances, see Moulton (1990). Therefore, the following equations are estimated by OLS:

$$\log w_{ij} = \beta_0 + \beta_2 X + \beta_j R_j + \varepsilon_{ij} \quad (5)$$

$$\beta_j = \lambda_0 + \lambda_1 \gamma + \upsilon_j \quad (6)$$

where $\log w_{ij}$ is the log (gross) hourly wage of worker i from region j and X_1 contains all the standard variables of the wage equation[7], R_j is a set of region dummies. Equation (5) was estimated once on all regions and once on the 95 (C)MSA's for which we have additional information on people per square mile and cost of living. We experimented in equation (6) with various extra controls. The results are presented in Table 1.

Table 1 about here

We can conclude from estimation (1) that our density index explains 45% of the cross regional wage variation that is not explained by the standard observable worker characteristics. The estimated value of λ_1 is highly statistically significant Workers living in an area with a γ of one standard deviation (0.19) left from the mean earn 7.4% more than people living in an area where γ is equal to its mean. In estimation (2) we include average PUMA size as an explanatory variable to test whether the potential aggregation bias that we discussed in section 2 plays a role. This is not the case. Its effect is not significant and λ remains almost the same. Next, we add dummies for the main regions: North Central/Mid West, South and West (North East is control). Wages in the North East turn out to be higher. Our density index only decreases slightly. In (5), we test whether controlling for people per square mile eliminates the effect of our density index.

[6] An additional problem of measuring the effect of labor market density on wages at the PUMA level is, as the referees pointed out, that the causality might be reversed. Firms in high productive regions could post higher wages to attract workers from other regions. In the CPS estimates, this mechanism is not relevant because there we use regions that consist of multiple PUMA's so that in general, both home and work location are in the same region. The mechanism described above, would then generate within region wage differentials while our estimates are based upon between region wage differentials.

[7] The explanatory variables are: a constant, female, unmarried, female*unmarried, black, 2 digit occupation and industry dummies, dummies for completed education (12, 14, 16, 18 years), education (yrs), cubic polinomial in experience and experience*education,female*experience, female* not married, female*not married* experience, $N = 66211$.

This turns out not to be the case. λ_1 drops but remains significant. Since we only have people-per-square-mile (ppsqm) data at the (C)MSA level, we also repeated (1) with the same regions as in regression (5) to check to what extent the fall in λ is due to fewer observations or to including ppsqm, see (4). It turns out that the inclusion of ppsqm makes λ_1 fall from -0.29 to -0.19. The drop in λ due to leaving the non-(C)MSA areas out is similar in magnitude: from -0.39 to -0.29. In (6) we included region dummies and λ_1 remains stable at -0.17 (3.11).

If, for the reasons we mentioned above, dense areas are attractive and if the stock of real estate is to some extent fixed than the real estate owners receive rents. We therefore, expect our index to be correlated with the cost of living. To see to what extent this is the case, we add the regional cost of living index of Dumond et al. (1999). This index is based on the American Chamber of Commerce Researchers Association (ACCRA) cost of living index for the period 1985:4 through 1995:2. In (7), (8) and (9) we see that this measure is also positively related to both our density index and people per square mile.

In (10), we estimate the impact of our density index at the PUMA level of aggregation with data from the 1% PUMS of the Census, see Ruggles and Sobek (1997).[8] The estimate of λ in the equivalent of (6) is -0.066 (16.76). Again, we conclude that workers in denser areas earn higher wages.[9]

5 Discussion

We have shown that we can give a meaningful structural labor market interpretation of the Ellison and Glaeser (1997) index of concentration. The large and significant effect that our density index has on wages suggests that it captures important effects. Our index is particularly useful for testing theories that predict that the scale of the labor market matters. From a theoretical point of view, our index is more attractive than the obvious alternative: people per square mile. This index is available for (C)MSA's only. For those regions for which it is available, our index and people per square mile are correlated. However, both are statistically significant in a wage equation. In Teulings and Gautier

[8] We have 224271 observations.

[9] When we aggregate up the Census regions to the CPS levels and place the same restrictions on the hours worked variable in the CPS, the Census estimate of the density index remains considerably smaller than the CPS one. The correlation between the sets of regional dummies obtained from (5), using the Census data and the CPS data is only 0.31.

(2002), we successfully apply our density index to analyze the effect of the efficiency of the search process on the distribution of workers and jobs across regions.

6 Literature

Burda M. S. Profit (1996), "Matching across space: Evidence on mobility in the Czech Republic", *Labour Economics*, 3, 255-278.

Coles, M. and E. Smith (1998), "Cross-Section estimation of the matching function: Evidence from England and Wales, *Economica*, 63, 589-598.

Diamond P.A. (1991), Aggregate demand management in search equilibrium, *Journal of Political Economy*, 89, 798-812.

Dumond, M., B.T. Hirsch and MacPherson (1999), Wage differentials across labor markets and workers: does cost of living matter? *Economic Inquiry*, 37, 577-598.

Ellison G. and E.L. Glaeser (1997), Geographic concentration in U.S. manufacturing industries: A dartboard approach, *Journal of Political Economy*, 105-5, 889-927.

Glaeser E.L and D.C. Mare (2001), Cities and Skills, *Journal of Labor Eonomics*, 19-2, 316-342.

Jaeger D.A., S. Loeb, S. Turner and J. Bound (1998), Coding geographic areas across census years: creating consistent definitions of metropolitan areas, *NBER working paper* 6772, Cambridge MA.

Lucas, R. (1988), On the mechanics of economic development, *Journal of Monetary economics*, 22, 3-42.

McFadden, D., (1973), Conditional logit analysis of qualitative choice behavior, in P. Zarembka ed., *Frontiers in Econometrics.* New York: Academic Press.

Moulton B.R. (1990), An illustration of a pitfall in estimating the effects of aggregate variables on micro units, *Review of Economics and Statistics*, 72, 334-338.

Roback J. (1982), Wages, rents and the quality of life, *Journal of Political Economy*, 90, 1257-1278.

Ruggles S. and M.Sobek, (1997), *Integrated Public Use Microdata Series*: Version 2.0, Minneapolis: University of Minnesota, 1997, Historical Census Projects,

Teulings C.N. and P.A. Gautier (2002), "Search and the city", *TI discusion Paper*, TI 2002-061/3, Tinbergen Institute, Amsterdam/Rotterdam, Wasmer E. and Y. Zenou (1999), Does space affect search? a theory of local unemployment, *CEPR Discussion Paper*, 2157.

Wheeler C.H. (2002), Search, sorting and urban agglomeration, *Journal of Labor Economics*, 19-4, 889-927.

A Appendix

A.1 Definitions

- **MSA** : relatively freestanding and not closely associated with other MA's, typically surrounded by non-metropolitan areas; the title of an MSA contains the name of its largest city and up to two additional city names.

- **CMSA**: consolidated metropolitan area. MA of more than 1 million people which may include one or more large urbanized counties that demonstrate very strong internal economic and social links within a CMSA. An example of a large CMSA is New York-New Jersey-Long Island.

B Tables and Figures

Table 1: Estimation results

		N	R^2	estimate	|t-value|
	Dependent variable: log hourly wage				
1	γ	150	0.454	-0.391	11.10
2	γ	150	0.484	-0.403	11.43
	avsize			6.109	1.41
3	γ	150	0.553	-0.382	11.31
	ncmw			-0.061	3.36
	south			-0.085	5.05
	west			-0.028	1.30
4	γ	95	0.242	-0.294	5.46
5	γ	95	0.415	-0.185	3.64
	ppsqm			0.096	4.12
6	γ	95	0.4388	-0.175	3.25
	ppsqm			0.083	3.56
	ncmw			-0.025	0.24
	south			-0.056	2.90
	west			0.004	0.90
	Dependent variable: log cost of living				
7	γ	95	0.212	-0.348	5.24
8	γ	95	0.531	-0.118	1.99
	ppsqm			0.024	8.03
9	γ	95	0.676	-0.087	1.74
	ppsqm			0.018	8.27
	ncmw			-0.082	4.19
	south			-0.089	4.93
	west			0.005	0.20
	Dependent variable: log hourly wage (Census)				
10	γ	1097	0.204	-0.066	16.76

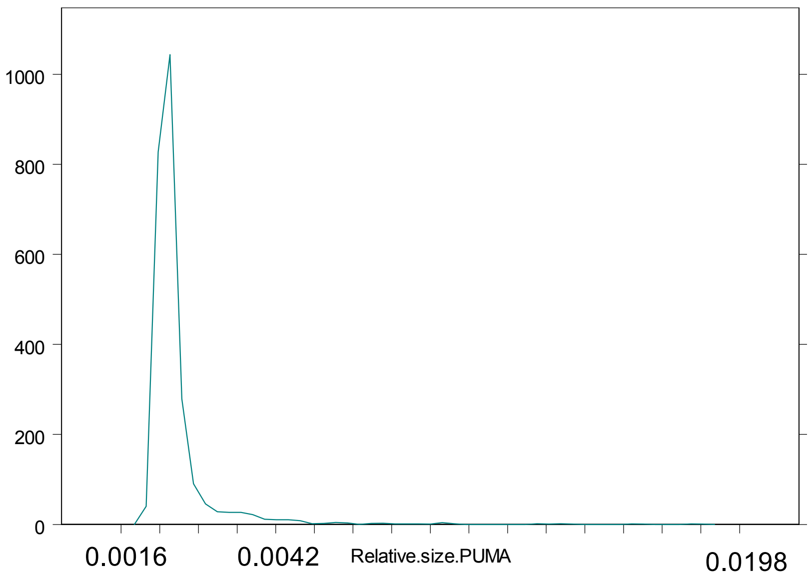

Figure 1: Density of area sizes, mean = 0.001, $\sigma^2 = 0.001$

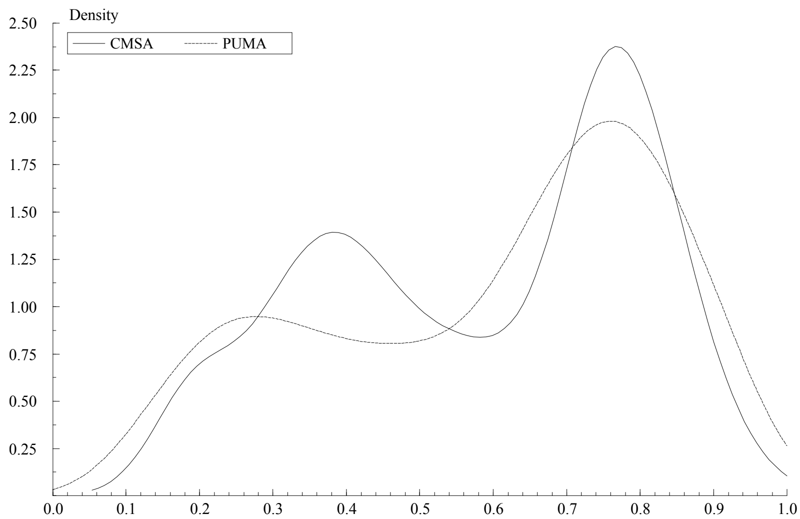

Figure 2:

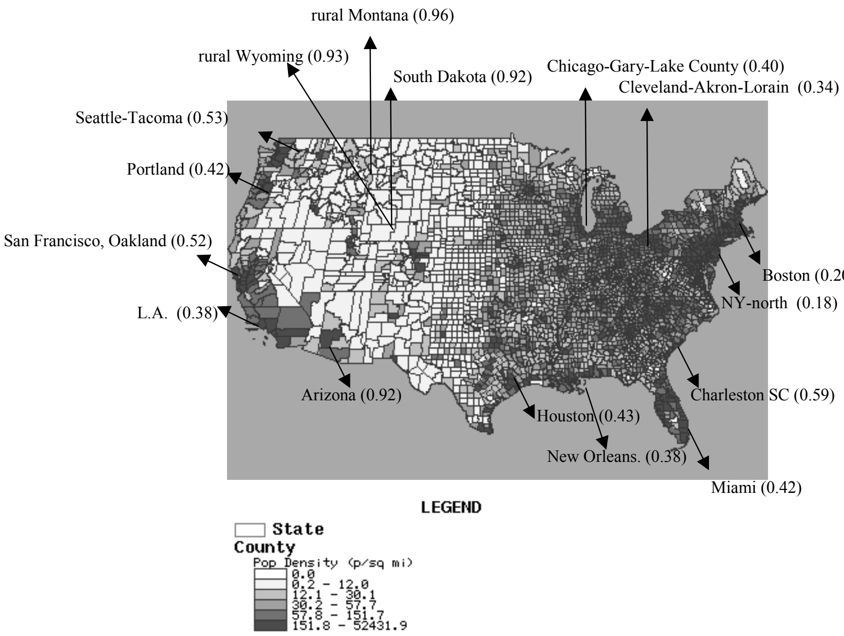

Figure 3: The relation between persons-per-square-mile and γ_{CPS}

C States/(C)MSA's ranked from dense to non dense

1 DC Washington 0.18201

2 New Jersey N.Y.-North. N.J.-Long Island 0.1842

3 Florida Orlando 0.19529

4 Massachusetts Boston-Lawernce-Salem-Lowell-Brockton 0.19993

5 Minnesota Minneapolis-St.Cloud 0.21236

6 Connecticut Hartford-New Britain-Middletown-Bristo l 0.26075

7 Connecticut N.Y.-North. N.J.-Long Island 0.26413

8 New Jersey Philadelphia-Wilmington-Trenton 0.28888

9 Texas Dallas-Fort Worth 0.30358

10 Colorado Denver-Boulder 0.30739

11 Massachusetts Worcester 0.31041

12 Connecticut New Haven-Meriden 0.31172

13 Michigan Detroit-Ann Arbor 0.3181

14 Rhode Island Providence-Pawtucket-Woonsocket 0.31855

15 Georgia Atlanta 0.32859

16 New Jersey Atlantic City 0.33552

17 New York Buffalo-Niagara Falls 0.3356

18 Ohio Cleveland-Akron-Lorain 0.34439

19 Virginia Richmond-Petersburg 0.34729

20 New York N.Y.-North. N.J.-Long Island 0.34776

21 Michigan Lansing-East Lansing 0.35209

22 Virginia Washington 0.36757

23 Louisiana Baton Rouge 0.37122

24 Tennessee Chattanooga 0.37835

25 New York Albany-Schenectady-Troy 0.37913

26 California Los Angeles city 0.37934

27 Louisiana New Orleans 0.38304

28 Massachusetts Springfield 0.38791

29 New York Syracuse 0.38889

30 Massachusetts Providence-Pawtucket-Woonsocket 0.39096

31 Kentucky Louisville 0.39682

32 Indiana Chicago-Gary-Lake County 0.39885

33 Maryland Baltimore 0.40908

34 Michigan Grand Rapids 0.41137

35 Tennessee Knoxville 0.41209

36 Illinois St. Louis 0.41564

37 Florida Miami-Fort Lauderdale 0.41688

38 Oregon Portland 0.41916

39 Illinois Chicago-Gary-Lake County 0.41966

40 Kentucky Cincinnati-Hamilton 0.42248

41 Missouri St. Louis 0.42903

42 Maryland Washington 0.43077

43 Texas Houston-Galveston-Brazoria 0.43306

44 Connecticut non (C)MSA area 0.43436

45 Virginia Norfolk-Virginia Beach-Newport News 0.43863

46 Michigan Flint 0.44093

47 Illinois Rockford 0.4426

48 North Carolina Fayetteville 0.44308

49 Connecticut New London-Norwich 0.44369

50 Pennsylvania Philadelphia-Wilmington-Trenton 0.44455

51 Kansas Kansas City 0.44475

52 North Carolina Greensboro-Winston-Salem-High Point 0.45378

53 California Sacramento 0.47082

54 California Modesto 0.47128

55 Tennessee Memphis 0.48094

56 Texas Beaumont-Port Arthur 0.48404

57 Ohio Cincinnati-Hamilton 0.48589

58 Florida Melbourne-Titusville-Palm Bay 0.49493

59 Washington Spokane 0.49705

60 Pennsylvania Harrisburg-Lebanon-Carlisle 0.49721

61 Missouri Kansas City 0.4975

62 Indiana Fort Wayne 0.49753

63 South Carolina Columbia 0.50242

64 California Fresno 0.50661

65 New York Rochester 0.50816

66 Texas Austin 0.51424

67 Iowa Des Moines 0.51545

68 California San Francisco-Oakland-San Jose 0.51671

69 California Bakersfield 0.52003

70 Washington Seattle-Tacoma 0.5322

71 Mississippi Jackson 0.53408

72 Indiana Indianapolis 0.53501

73 Wisconsin Madison 0.53776

74 Tennessee Nashville 0.54252

75 Oregon Eugene-Springfield 0.54305

76 Illinois Peoria 0.54461

77 Pennsylvania Allentown-Bethlehem 0.55004

78 Massachusetts non (C)MSA area 0.5597

79 Kentucky Lexington-Fayette 0.56318

80 Illinois Davenport-Rock Island-Moline 0.56678

81 Oklahoma Oklahoma City 0.5766

82 Georgia Macon-Warner Robins 0.57871

83 Ohio Youngstown-Warren 0.58012

84 Nevada Reno 0.58201

85 Ohio Dayton-Springfield 0.58718

86 Nebraska Omaha 0.59182

87 Wisconsin Milwaukee-Racine 0.59276

88 South Carolina Charleston 0.59322

89 North Carolina Raleigh-Durham 0.59538

90 Colorado Colorado Springs 0.59608

91 Texas San Antonio 0.59681

92 Wisconsin Appleton-Oshkosh-Neenah 0.60155

93 North Carolina Charlotte-Gastonia-Rock Hill 0.60192

94 New Hampshire non (C)MSA area 0.60363

95 California Salinas-Seaside-Monterey 0.61512

96 Ohio Toledo 0.61773

97 Indiana Louisville 0.61824

98 Florida Fort Pierce 0.62039

99 Alabama Birmingham 0.62628

100 Alabama Montgomery 0.62861

101 Tennessee Johnson City-Kingsport-Bristol 0.62942

102 South Carolina non (C)MSA area 0.64862

103 Michigan Saginaw-Bay City-Midland 0.64866

104 Ohio Columbus 0.66128

105 South Carolina Greenville-Spartanburg 0.66141

106 Pennsylvania Pittsburgh-Beaver Valley 0.67295

107 Georgia non (C)MSA area 0.67496

108 Ohio non (C)MSA area 0.67521

109 Oklahoma Tulsa 0.67564

110 Indiana non (C)MSA area 0.6877

111 Delaware non (C)MSA area 0.69159

112 Maryland non (C)MSA area 0.69311

113 New York Binghampton 0.69981

114 Utah Salt City-Ogden 0.71502

115 Vermont non (C)MSA area 0.71874

116 Florida Jacksonville 0.71967

117 Indiana Evansville 0.72104

118 Pennsylvania York 0.7232

119 Pennsylvania Scranton-Wilkes-Barre 0.72731

120 Mississippi non (C)MSA area 0.72741

121 Virginia non (C)MSA area 0.72983

122 West Verginia non (C)MSA area 0.73126

123 Alabama non (C)MSA area 0.73288

124 Georgia Augusta 0.73401

125 Iowa non (C)MSA area 0.73448

126 New York Utica-Rome 0.73627

127 North Carolina non (C)MSA area 0.74169

128 Maine non (C)MSA area 0.7418

129 Arkansas Little Rock-North Little Rock 0.74541

130 Kentucky non (C)MSA area 0.7466

131 Virginia Johnson City-Kingsport-Bristol 0.74663

132 Tennessee non (C)MSA area 0.75257

133 Michigan non (C)MSA area 0.75505

134 New York non (C)MSA area 0.75518

135 Illinois non (C)MSA area 0.75701

136 Florida Tampa-St. Petersburg-Clearwater 0.76119

137 Texas Killeen-Temple 0.76615

138 Wisconsin non (C)MSA area 0.76934

139 Texas Corpus Christi 0.76942

140 Arkansas non (C)MSA area 0.77266

141 Washington non (C)MSA area 0.77861

142 Missouri non (C)MSA area 0.7907

143 Minnesota non (C)MSA area 0.79186

144 Pennsylvania non (C)MSA area 0.794

145 Alabama Mobile 0.80454

146 Louisiana non (C)MSA area 0.80639

147 Kansas Wichita 0.80993

148 New Mexico non (C)MSA area 0.82036

149 Florida non (C)MSA area 0.82273

150 North Dakota non (C)MSA area 0.82542

151 Colorado non (C)MSA area 0.85206

152 California non (C)MSA area 0.86129

153 Nebraska non (C)MSA area 0.86141

154 Idaho non (C)MSA area 0.86383

155 Kansas non (C)MSA area 0.86384

156 Oklahoma non (C)MSA area 0.87242

157 Oregon non (C)MSA area 0.87992

158 Utah non (C)MSA area 0.8916

159 Texas non (C)MSA area 0.89445

160 Arizona non (C)MSA area 0.91724

161 South Dakota non (C)MSA area 0.91771

162 Nevada non (C)MSA area 0.92344

163 Wyoming non (C)MSA area 0.93019

164 Montana non (C)MSA area 0.9479