# The Thick Market Effect on Local Unemployment Rate Fluctuations *

Li Gan† and Qinghua Zhang‡

March 2003

**Abstract**

This paper presents a model studying the effect of city size on a city's unemployment rate. The model demonstrates that due to the thick market effect on improving the matching probability between jobs and workers, larger cities have lower unemployment rates, shorter unemployment cycles, and shallower recessions. Our empirical tests are consistent with the predictions of the model. In particular, we find that an increase of two standard deviations in city size lowers the unemployment rate by about a half percentage point, decreases peak unemployment rate by .3 percentage points, and reduces the unemployment cycles by about one month.

Key words: thick market effect, local unemployment fluctuation, matching.

*We thank Vernon Henderson, John Driscoll, Bent Sorensen, and Peter Hansen for their help during the writing of an earlier version of this paper. We also would like to thank Jingyuan Yin for research assistance. All remaining errors are ours.

†Department of Economics, University of Texas, Austin, TX 78712. gan@eco.utexas.edu

‡Department of Economics, University of Texas, Austin, TX 78712. zhangq@eco.utexas.edu

# 1 Introduction

Unemployment rates vary widely across cities in the United States. Among the 295 Primary Metropolitan Statistical Areas (PMSAs), the average unemployment rate from 1981 to 1997 ranged from 2.4% in Columbia, Missouri (PMSA code 1740), to 19.6% in McAllen-Edinburg-Mission, Texas (PMSA code 4880). There are two common hypotheses to explain this phenomenon in the literature: the industry composition hypothesis and the risk diversification hypothesis.

The industry composition hypothesis is very intuitive: different cities have different industry compositions.[1] Thus, nation-wide industry-specific shocks will have different composite effects on unemployment rates in different cities. The second hypothesis is based on the observation that in the local labor market, prosperous industries absorb the unemployment of those experiencing contractions. Therefore, a city with a more diversified industry structure has a lower variance in the labor demand. As a result, the frictional unemployment rate in this city is also lower. For example, Mills and Hamilton (1984) argue that a larger city is usually more industrially diverse and thus has a lower unemployment rate. Neumann and Topel (1991) provide a formal model on the effect of risk diversification.

A few empirical studies have confirmed both the industry composition hypothesis and the risk diversification hypothesis, e.g., Simon (1988) and Neumann and Topel (1991). Simon's study is based on data at the 2-digit SIC level for 91 large PMSAs of the U.S. over 1977–1981. He finds that the frictional unemployment rate declines as local industrial diversity rises. He defines the frictional unemployment rate as the city's aggregate unemployment rate net of the effects of national shocks and industry composition. Using U.S. data at the state level over 1950–1985, Neumann and Topel demonstrate that after the effect of industry composition is controlled for, the unemployment rate is significantly and persistently lower in labor markets where the sectoral demand risk is more diversified.

This paper offers a model that provides yet another explanation of the wide difference

[1] In this paper, the term "city" has the same meaning as the the term "PMSA."

in unemployment rates across cities: the thick market effect and the agglomeration economy associated with it. In addition to the difference in the average unemployment rate, our model can also explain the variation in the frequency/duration, and the peak unemployment rate fluctuations.

The intuition of our model is as follows. When number of unemployed workers and job openings both increase, the matching probability rises due to the effect of thick market. When market is thick enough to reach a certain size, workers expected returns from search is higher than the cost of search, workers start to search and matches occur. Due to the thick market effect, unemployed workers accumulate in the city until the local labor market reaches the certain size. The larger a city's local market, the faster for the city to reach the certain size.[2] Therefore, the local labor market becomes active at a certain frequency, which results in cyclical fluctuations of the city's unemployment rate. For example, job fairs in a city are usually held at intervals instead of continuously.

In particular, the model predicts a certain type of agglomeration economy, that is, larger cities on average have shorter unemployment cycles and shallower recessions. Because a larger city typically generates more unemployed workers during each time period, it takes less time for the city's labor market to reach a large enough size. Therefore, its labor market becomes active more frequently and its unemployment cycles are shorter on average. It also follows that the peak unemployment rate and average unemployment rate are both lower. A lower peak unemployment rate indicates a shallower recession. This particular type of agglomeration economy is based on the thick market effect in the local labor market.[3]

Empirically, this paper tests three predictions of our model: (1) In addition to the effects

[2]Diamond (1982) presents a model of the thick market effect that hinges on the search cost, instead of on the matching probability as in our model. His idea is that the more activity there is on one side of the market, the lower the contacting costs faced by those who are looking for trading partners are on the other side. Howitt and McAfee (1987) provide an explicit model of the labor market, in which Diamond's thick market effect is present.

[3]For a general discussion of the agglomeration economy, see Henderson (1986, 1988). In addition, Wilson (1988) provides an empirical test for the agglomeration economy.

of risk diversification and industry composition, the unemployment rate in a city should be negatively correlated with city size. (2) The length of an unemployment cycle is shorter in a larger city size. (3) The peak unemployment rate drops as city size increases.

While studying how the average and peak unemployment rates vary across cities has important policy implications, it is equally important to understand the variation in the duration of the unemployment cycle. The duration of an unemployment cycle is the time length of the cycle.[4] Although the length of unemployment cycles is different from the mean unemployment duration of individuals, they are positively related. In a simple version of model, a cycle of unemployment starts with full employment. The number of unemployed workers accumulates over time until it reaches a critical size when matches occur and everyone is employed. In this model, if the accumulation of unemployment is linear, then the length of the unemployment cycle is twice as long as the mean unemployment duration of individuals.

To find the relationship between the unemployment rate and city size, we use a linear regression model that includes the log of average city size as one of the explanatory variables. One way to test the negative correlation between the length of an unemployment cycle and city size is by using spectral analysis. If we think of the time series as compounded cycles with different frequencies, the spectral density of a certain frequency measures how much the cycle associated with this specific frequency contributes to fluctuations in the time series. We consider two types of frequencies: the max-frequency and the mean-frequency. Since a frequency is the inverse of a cycle length, our model predicts that the two types of frequencies are positively correlated with the city size.

Another way to find out the relationship between length of unemployment cycles and city size is by conducting a duration analysis in which we decompose a whole unemployment cycle into two stages: the peak-to-trough stage (the expansion in the economy) and the trough-to-peak stage (the contraction in the economy). In particular, our model predicts that the length of the trough-to-peak is negatively related to the city size.

Testing a negative relationship between peak unemployment and city size is relatively

[4] A formal definition of the duration of an unemployment cycle is given in Section 5.

straightforward. After identifying peak points in unemployment cycles, we construct an average peak unemployment rate for each city and then find out its relationship with the log of average city size.

The empirical results in this paper are consistent with the previous three predictions of the model. In particular, we find that an increase of two standard deviations in city size lowers the unemployment rate by about a half percentage point, reduces the unemployment cycle by about one month, and lowers the peak unemployment rate by .3 percentage points.

The rest of the paper is organized as follows: Section 2.1 presents the theoretical model that underlies the later empirical tests. Section 2.2 discusses the data. Section 3 investigates how the level of the unemployment rate is influenced by city size. Section 4 conducts the spectral analysis on patterns of cyclical fluctuations in unemployment rate. Section 5 carries out the duration analysis on the average length of unemployment cycles; it also studies the peak unemployment rates. Section 6 concludes.

# 2 The Model and the Data

## 2.1 The Model

In this section, we present a simple model that illustrates the effect of the thick market and its associated agglomeration economy on local unemployment rate fluctuations.

Let $N$ be the number of workers in a city who are immobile across cities, and let this be the measure of city size. Let $U$ be the number of unemployed workers in the city, and let $V$ be the job openings in the city. Both workers and jobs are heterogeneous, denoted as $a$ for a worker and $b$ for a job. One may interpret $a$ as an index for ability and $b$ as the capital stock invested in the job. For illustration purpose, assume each firm has only one opening.

A firm's profit from a job opening $b$ is given by:

$$\pi(a,b) > 0 \text{ if and only if } \; a > b. \tag{1}$$

where $\partial\pi(a,b)/\partial a > 0$. The firm prefers a higher $a$ to a lower $a$. Equation (1) also indicates a match occurs if and only if $a > b$. Intuitively, a certain ability is required for a match to be productive.

An unemployed worker chooses whether to actively look for job or not. If he does, he incurs certain search cost, $c(a)$; otherwise, his utility is zero. If the worker participates the job market, his utility function is given by:

$$W(a,b) = \begin{cases} w(a,b) - c(a), & \text{if} \quad \text{matches with firm b;} \\ -c(a), & \text{if} \quad \text{no match.} \end{cases} \tag{2}$$

where $w(a,b)$ is the wage from this match. We assume $w(a,b) > c(a)$ and $\partial W(a,b)/\partial b = \partial w(a,b)/\partial b > 0$. since a more capital intensive job typically pays more. The worker will accept any job offer, but he prefers a more capital intensive job. The worker is uncertain if he is going to get any offer before he starts actively searching for jobs..

The matching mechanism considered here is very simple. All job openings $b$'s are posted in a centralized bulletin. All workers also list their ability $a$ on this bulletin. Firm $b$ is willing to match with any $a$ if $a > b$ but prefers a higher $a$ to a lower $a$. Worker $a$ is willing to match with any firm $b$ but prefers a higher $b$ to a lower $b$. A match occurs between $a$ and $b$ if and only if that $b$ is the best choice available for $a$ and vice versa. In such a model, Gan and Li (2002) show that the probability of matching increases when total number of workers and total of number jobs increase. To illustrate the intuition of this result, consider the following example.

First, consider a market with one worker and one job opening. Let $a$ and $b$ are randomly drawn from the same distribution. The matching probability in this case is 1/2 since $\Pr(a > b) = 1/2$.

Then consider a market with two workers, $(a_1, a_2)$ and two openings, $(b_1, b_2)$. All $(a_1, a_2)$ and $(b_1, b_2)$ are randomly drawn from the same distribution. Further let $a_{(1)} > a_{(2)}$ and $b_{(1)} > b_{(2)}$. Since $a_i$ and $b_i$ are from the same distribution, the order statistics $a_{(i)}$ and $b_{(i)}$ are also from the same distribution. Given this, we have:

$$\Pr(a_{(1)} > b_{(1)}) = 1/2, \quad \text{and} \quad \Pr(a_{(2)} > b_{(2)}) = 1/2. \tag{3}$$

If (3) were the only case that workers and openings match, the matching probability is 1/2.

However, an additional chance exist when $a_{(1)} < b_{(1)}$ and $a_{(2)} < b_{(2)}$ since it is still possible to have $a_{(1)} > b_{(2)}$. This additional chance of matching is the source of the effect of a thicker market. In fact, the matching probability in this case is 7/12.

Formally, Gan and Li (2002) present mathematical matching probabilities as both number of workers and number of openings increase. They conclude that a thicker market has a higher matching probability. This result holds when number of workers does not equal to number of workers $U \neq V$, as long as $U/V$ is bounded from infinity or zero as $U \to \infty$. In the following discussion we let $U = V$ for simplicity.

Let a worker $a$'s probability of being matched be $\Pr(a, U)$. The thick market effect indicates that $\partial \Pr(a, U)/\partial U > 0$. The expected return from participating labor market is:

$$\begin{aligned} E(W) &= \int \left(\Pr(a,U)w(a,b) - c(a)\right) f_b(b)db \\ &= \Pr(a,U)w^*(a) - c(a) \end{aligned} \tag{4}$$

where $f_b(b)$ is the distribution of $b$ and $w^*(a) = E_b[w(a, b)]$.

When the number of unemployed $U$ (and number of openings $V$) increases, $\Pr(a, U)$ increases. When market size is large enough, $E(W)$ will become positive and the worker will participate in the job market. Let the critical size of the job market, $\bar{n}$, be such that:

$$E(W) = \Pr(a, \bar{n})w^*(a) - c(a) = 0. \tag{5}$$

The solution to (5), $\bar{n}(a)$, is a function of $a$. If $E(W)$ in (4) can be separated into two parts, i.e., $E(W) = P(U, V) \cdot s(a)$, the critical market size does not depend on the $a$, i.e., $\bar{n}(a) = \bar{n}$. In this case, all workers in the market assume the same critical market size.

More generally, we focus on the mean value of $\bar{n}(a)$. Let $\bar{n}^* = E[\bar{n}(a)]$. We claim that the the existence of such a critical minimum market size leads to the cyclical fluctuations in the unemployment rate in the local market.[5]

Let us normalize the time of a clearance of the local labor market as time $t = 0$. Then at the beginning of time $t = 1$ the unemployment in the local market is zero. Let $U_t$ be the

[5] A strategic version of the current model can be found in Zhang (2002).

number of accumulated unemployed workers by $t$. Let $T$ be the number of time intervals such that:

$$U_T \geq \bar{n}^* > U_{T-1}. \tag{6}$$

The inequalities in (6) say that $T$ is the smallest number of time intervals such that the accumulated number of unemployed workers in the local market will be larger or equal to the minimum market size $\bar{n}^*$. Assume that the separation rate of a worker-job pair during any time period is $\nu$, we have:

$$U_t = U_{t-1} + \nu(N - U_{t-1}), \quad 1 \leq t \leq T, \quad U_0 = 0.$$

Solving the above difference equation, we get:

$$U_t = N(1 - (1-\nu)^t), \quad 1 \leq t \leq T, \quad U_0 = 0. \tag{7}$$

The unemployment rate at the end of time $t$, denoted as $u_t$, is thus:

$$u_t = U_t/N = 1 - (1-\nu)^t, \quad 1 \leq t \leq T, \quad U_0 = 0. \tag{8}$$

Clearly, $u_t$ increases with $t$, which reflects the fact that over time, as the unemployed workers accumulate in the local market, the unemployment rate goes up. The average unemployment rate over the time interval $[1, t]$ is:

$$\begin{aligned} \bar{u}_t &\equiv \frac{\sum_{i=1}^{t} u_i}{t}, \quad 1 \leq t \leq T, \quad U_0 = 0, \\ &= 1 - \frac{1-\nu}{\nu}\left(\frac{1-(1-\nu)^t}{t}\right). \end{aligned} \tag{9}$$

From (9), $\partial \bar{u}_t / \partial t > 0$. The logic is: since $u_t$ increases as $t$ increases, its average over $t$, $\bar{u}_t$, also goes up with $t$.

At time $T$, the number of accumulated unemployed workers just reaches the critical minimum size for the labor market to clear. According to (6) and (7),

$$N(1-(1-\nu)^T) \geq \bar{n}^* > N(1-(1-\nu)^{T-1}).$$

Rearrange the above inequality as follows:

$$T \geq \frac{\ln(1 - \bar{n}^*/N)}{\ln(1 - \nu)} > T - 1. \tag{10}$$

From (10), we can see that $T$ decreases as $N$ increases. Intuitively, it takes less time for a larger city to accumulate enough unemployed workers in the local labor market, given $\nu$. Because $T$ measures the length of time from the trough to the peak of an unemployment cycle, the length of unemployment cycles is therefore negatively correlated with city size.

At time $T$, the unemployment rate is at its highest. From (8), the peak point unemployment rate is given by:

$$u_T = 1 - (1 - \nu)^T. \tag{11}$$

Equation (11) says that the peak unemployment rate $u_T$ increases as $T$ increases. To relate with city size, the peak unemployment rate decreases as city size increases.

According to (9), the average unemployment rate over the time interval $[1, T]$ is:

$$\bar{u}_T = 1 - \frac{1 - \nu}{\nu} \left( \frac{1 - (1 - \nu)^T}{T} \right). \tag{12}$$

Again, Equation (12) says that the average unemployment rate over a cycle increases as $T$ increases. In terms of city size, our model predicts that the average unemployment rate over a cycle is lower for larger cities.

To better illustrate our model, we draw the unemployment fluctuation rate in two hypothetical markets in Figure 1. We let the probability of separation be constant at $\nu = .015$. The critical size of the market $\bar{n}^* = 5,000$. In the top graph in Figure 1, city size is 60,000. In the bottom graph in Figure 1, city size is 30,000. From the two graphs, we see it takes longer time for the smaller city to reach the critical size. The length of the cycle in the larger city is 5.75 while the length of the cycle in the smaller city is 12. The average unemployment rate in the larger city is about .05, while the average unemployment rate in the smaller city is about .10. By design, the peak unemployment rate in the larger city is 8.3%, while the smaller city's peak unemployment rate is 16.7%.

In summary, the model has three testable predictions: (1) The unemployment rate in a city should be negatively correlated with city size. (2) The length of unemployment cycles is shorter in a larger city. (3) Larger cities have lower peak unemployment rates.

### 2.2 The Data

The empirical analysis utilizes a sample of 295 PMSAs in the U.S. over the years 1981–1997. During this period, the U.S. economy experienced both recession and expansion.

The data on monthly unemployment rates is collected from the Employment and Earnings published by the Department of Labor's Bureau of Labor Statistics (BLS). Let the city $c$'s unemployment rate at time $t$ be $unempr_{ct}$. The employment data by PMSA is compiled from County Business Patterns, by summing up the city's employment over industries. Let us denote city $c$'s employment at time $t$ be $emp_{ct}$.

The industry employment information is obtained from the data that covers 543 industries at the 3-digit SIC level. We use the yearly employment data in County Business Patterns to calculate industry shares for each PMSA. We also use increments in the national employment by industry to approximate the nationwide industry-specific shock. The data on national employment by industry is obtained from the Bureau of Labor Statistics. Let $s_{ict}$ denote the employment share of industry $i$ in city $c$ at time $t$. Let $\triangle_{it}$ denote the nationwide employment growth rate of industry $i$ during time $t$. The industry composition effect on city $c$ at time $t$ will then be:

$$\begin{aligned} INDCOM_{ct} &= \textstyle\sum_i^{543} s_{ict} \times \triangle_{it}, \\ \text{where } c &= 1, 2, ..., 295, \\ i &= 1, 2, ..., 543, \\ \text{and, } t &= 1981:1, 1982:2, ..., 1997:12. \end{aligned} \tag{13}$$

Note here $s_{ict}$'s are the same for all the $t$'s in the same year, because for each city, its industry shares do not change much over months within a year.

The other variable is the risk diversification effect, denoted as $RISK_{ct}$. This variable measures uncertainty local labor demand that depends on the covariance of local labor demand across industries. Following Neumann and Topel (1991), we compile a variable $RISK$:

$$RISK_{ct} = s'_{ct}\Omega s_{ct}, \tag{14}$$

where $s_{ct}$ is the vector of industry employment shares of city $c$ at time $t$ and $\Omega$ is the covariance of nationwide industry-specific (detrended) shocks. The higher $RISK_{ct}$, the higher the uncertainty in local labor demand. Because the market friction tends to be greater when the uncertainty of the employment is higher, $RISK_{ct}$ affects the unemployment rate in a positive way.

Table 1 is a summary of statistics of the variables involved in the analysis of this paper. The sample period is January 1981 – December 1997. The unemployment rate, *unempr*, is measured in percentage points. The size in Table 1 is the city's total labor force. Since variable log($size$) will be used, we list the summary statistics of the log of the average city size. Unemployment benefits, denoted as *benefit*, are another important factor affecting unemployment rates. We use the ratio of average weekly benefit to average weekly total wage to represent unemployment benefits. The state-by-state ratio is obtained from the U.S. Department of Labor (http://www.doleta.gov). If a PMSA is across more than one state, we assign the mean ratio of these states to the PMSA. The national unemployment rate, *nunempr*, is calculated from our sample.

Table 1 shows that both city size and unemployment rate vary significantly. The average labor force ranges from 247,289 in Enid, Oklahoma (PMSA code 2340), to 3,532,300 in Los Angeles-Long Beach, California (PMSA code 4480). The average unemployment rate ranges from 2.4% in Columbia, Missouri (PMSA code 1740) to 19.6% in McAllen-Edinburg-Mission, Texas (PMSA code 4880).

In Figure 2, we draw mean unemployment rates and log of city size. The straightline in the figure is the fitted line. The slope of the fitted line is -.366 (.123). To ease the potential concern about "outliers," we delete cities that have unemployment rates larger than 15%. The fitted slope (not shown in the figure) is still significantly negative at -.250 (.107). In the following section, we will investigate this relationship in more detail.

Table 1: Summary Statistics of PMSA Averages (1981–1997)

| | mean | standard deviation | minimum | maximum | original data frequency |
|---|---|---|---|---|---|
| PMSA average | | | | | |
| unempr | 6.60 | 2.30 | 2.43 | 19.58 | monthly |
| emp | 232,110 | 397,482 | 19,931 | 3,262,702 | monthly |
| size | 247,289 | 425,347 | 19,931 | 3,532,300 | monthly |
| log($size$) | 11.71 | 1.07 | 9.94 | 15.7 | monthly |
| benefit | .362 | .047 | .264 | .470 | yearly |
| INDCOM | .00208 | 0.000506 | .207 | .393 | monthly |
| RISK | 0.00266 | 0.00134 | 0.000979 | 0.0130 | yearly |
| INDCOM × RISK | 8.2894E-6 | 6.938E-6 | 1.575E-6 | .0000683 | monthly |
| National average | | | | | |
| nunempr | 6.18 | 1.29 | 4.01 | 10.52 | monthly |

## 3 City Size and Mean Unemployment Rate

In this section, we examine the relationship between the average level of the unemployment rate and city size. The basic model we are interested in is as follows:

$$unempr_{ct} = \alpha_c + Z(t) + X_{ct}\beta + \eta \log(size_c) + \epsilon_{ct}, \tag{15}$$

where the $X_{ct}$ is a vector of control variables. In particular, we consider:

$$X_{ct} = \{RISK_{ct}, INDCOM_{ct}, RISK_{ct} \times INDCOM_{ct}, benefit_{ct}\}, \tag{16}$$

where the variable $RISK_{ct}$ is constructed in (14), and the variable $INDCOM_{ct}$ is constructed in (13). The expected sign for $RISK_{ct}$ is positive and for $INDCOM_{ct}$ is negative. The coefficient for the interaction term is unclear.

To investigate the relationship between the unemployment rate and city size, we include an additional term $\log(size_c)$ in the model. City size is defined by the city's average total labor force in our sample period. The coefficient on the log of average city size, $\eta$, is expected to be negative: the larger the city size, the lower the unemployment rate.

In Equation (15), the term $Z(t)$ is used to control the effect of national business cycles. We consider two alternative specifications of $Z(t)$. First, we let $Z(t) = Z_t$. This is a model with a fixed time effect. Second, we let:

$$Z(t) = \{nunempr_t, t, t^2, t^3\}, \tag{17}$$

where the time trend $t$ is calculated by (year-1981) * 12 + month, and $nunempr_t$ is the national unemployment rate at time $t$. Since a third order polynomial is included in (17), the second specification is reasonably flexible to control any potential aggregate time effect. $Z(t)$ in (17) will be used again in the spectral analysis in Section 4 to control for the time trend.

Another term $\alpha_c$ in (15) represents the unobserved city heterogeneity. Since the variable $\log(size_c)$ does not change over time, we cannot use a fixed city effect model. Instead, we let $\alpha_c$ be a random variable, such that $E(\alpha_c|X_{it}, \log(size_c)) = 0$. This specification represents a random city effect model. For comparison purposes, we estimate models that do not include the term $\log(size_c)$.

Table 2 lists the regression results from alternative specifications of (15). The first two columns are the estimation results from a time fixed-effect model, and the last two columns list results from a model that uses $Z(t)$ in (17) to control for the time effect.

Column (1) and Column (3) do not have the thick market effect, while Column (2) and Column (4) include the thick market effect. In all four specifications, the coefficients for the variable $INDCOM$ are significantly negative, and the coefficients for the variable $RISK$ are significantly positive, as predicted. These results support the two previous hypotheses of local unemployment: the industry composition hypothesis and the risk diversification hypothesis.

More importantly, in the regression results reported in Columns (2) and (4), the log of city size has a significantly negative effect on a city's unemployment rate. In Column (2), where the fixed time effect is used, the coefficient for the $\log(size)$ is -.237 (.117). In Column (4), where the time trend and national unemployment rate are used, the coefficient for the $\log(size)$ is -.292 (.117). The first prediction of our model is supported: a larger city has a lower unemployment rate.

Table 2: Unemployment Rate Mean Regression Results

| Variables | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| time fixed effect | yes | yes | no | no |
| city random effect | yes | yes | yes | yes |
| INDCOM | -28.3 | -28.3 | -8.92 | -8.92 |
| | (2.01) | (2.01) | (.966) | (.966) |
| RISK | 231.4 | 229.6 | 163.9 | 161.6 |
| | (13.7) | (13.7) | (13.6) | (13.6) |
| INDCOM $\times$ RISK | -371.7 | -358.2 | -894.8 | -879.7 |
| | (323.5) | (323.5) | (295.8) | (295.9) |
| unemployment benefit | 10.42 | 10.40 | 10.51 | 10.49 |
| | (.286) | (.285) | (.282) | (.282) |
| national unempr | | | .910 | .910 |
| | | | (.0075) | (.0075) |
| (time trend/100) | | | 2.62 | 2.62 |
| | | | (.178) | (.178) |
| $(\text{time trend}/100)^2$ | | | -3.38 | -3.38 |
| | | | (.164) | (.164) |
| $(\text{time trend}/100)^3$ | | | 1.00 | 1.00 |
| | | | (.046) | (.046) |
| log(mean labor force) | | -.237 | | -.292 |
| | | (.117) | | (.117) |
| $R^2$ | .261 | .275 | .153 | .162 |
| No. of Obs. | 51274 | 51274 | 51274 | 51274 |

To compare the magnitude of the effects of all three hypotheses, we calculate the changes in the unemployment rate given an increase of two standard deviations on each of the three variables $INDCOM$, $RISK$, and $\log(size)$. If we apply the estimates from Column (4), the unemployment rate would decrease by .015 percentage points if $INDCOM$ increases, increase by .43 percentage points if $RISK$ increases, and decrease by .60 percentage points if city size increases. The effect of the thick market is significant and has roughly the same magnitude as the effect of risk diversification.

## 4 City Size and the Frequency of Unemployment Fluctuations — a Spectral Analysis

Our focus in this section is on the relationship between the frequency of fluctuations in a city's unemployment rate and its size.

We conducted the spectral analysis on three samples. The first sample consists of 139 PMSAs in the U.S. during 1981–1997. The unemployment rate data is monthly. For each PMSA, the number of observations of the unemployment rate is at least 200, indicating at most 4 missing values in the monthly unemployment rate. The second sample contains 168 PMSAs in the U.S. during 1983–1997. For each PMSA, the number of observations of the unemployment rate is at least 176, again indicating at most 4 missing values. As for the third sample, the period is 1986–1997; and there are 204 PMSAs. For each PMSA, the number of observations of the unemployment rate is at least 140. Because we allow for each PMSAs at most 4 missing values in the monthly unemployment rate, as the sample period becomes longer, there are fewer qualified PMSAs remaining in the sample. We can see that any one of the three sample periods experienced both recession and expansion in the U.S. economy. Each sample contains PMSAs of all sizes. Since the spectral analysis on the three samples all show similar results, for convenience, in this paper we only present the regression results based on the first sample; that is, the one with the longest sample period (i.e., January 1981–December 1997).

We run regressions of the unemployment rates $unempr_{ct}$ on $X_{ct}$ in (16) to to control for the effect of industry composition and risk diversification, and on $Z(t)$ in (17) to control for the effect of the time trend. Different from the regression in (15), here the regression is conducted city by city.[6] Our objective in this section is to conduct the spectral analysis in the frequency domain of the residuals city by city.

### 4.1 The Band Spectrum Regression and Filtering

The regression to be carried out here is called a band spectrum regression. It is conducted in the frequency domain. Since we want to examine the frequency of the unemployment rate, it is natural to use a regression in the frequency domain to control for the effects of the trend variables on the frequency of the unemployment rate. The band spectrum regression method adopted here follows Corbae, Ouliaris and Phillips (2002).

We divide the frequency domain into three bands. Band 1 consists of frequencies that correspond to cycles with a length from 2 to 4 months. This is a high frequency band. Band 2 includes frequencies associated with cycles longer than 4 months but shorter than 18 months. This is a medium frequency band. Since a typical waiting period in the job search process falls within this band, studying this band may reveal important information on the average waiting period in the job search process.[7] Band 3 is a low frequency band, consisting of frequencies corresponding to cycles longer than 18 months. This band includes the national business cycle frequencies, since according to National Bureau of Economic Research definitions, a business cycle in the U.S. at the national level has a length of between 18 and 96 months.

Let $W$ denote a discrete Fourier transformation such that for any time series $y$ of length $T$, $W$ is a $T \times T$ matrix and $Wy$ is the discrete Fourier transformation of $y$. The $T$ fundamental frequencies in $Wy$ are $0, 2\pi/T, 4\pi/T, ..., 2\pi(T-1)/T$. Let $A^j$ be a $T \times T$ diagonal matrix with value 1 at the $k$-th row if $2\pi(k-1)/T$ lies within Band $j$ as previously defined, and which

[6] Therefore, a fixed time effect is not applicable here.

[7] The mean unemployment duration of individuals is 3.8 months during 1994–2000 (Abraham and Shimer, 2001). According to our discussion in Section 1, the length of an unemployment cycle is roughly twice as long as the mean unemployment duration.

otherwise has a value of 0. In other words, by taking the product of $A^j$ and $Wy$, we can zero out all the fundamental frequencies in $Wy$ that lie outside of Band $j$.

The regression model specifies:

$$\begin{aligned} Wunempr_{ct} &= A^1WZ\alpha_c^1 + A^2WZ\alpha_c^2 + A^3WZ\alpha_c^3 \\ & \quad A^1WX_c\beta_c^1 + A^2WX_c\beta_c^2 + A^3WX_c\beta_c^3 + Wu_c \end{aligned} \tag{18}$$

where $\alpha_c^i, \beta_c^i$, and $i = 1, 2, 3$ are parameters to be estimated and which vary by city. Note that Equation (18) allows parameters to be different in different bands, capturing the possibility that the relationship between unemployment rates and control variables is frequency-dependent. After the regression, we take the residuals for each city $c$ and conduct the inverse Fourier transformation. The resulting time series is an estimate of the detrended unemployment rate, denoted $\{u_{ct}\}$.

Before we conduct the spectral analysis, there is one more step to go. We need to smooth the irregular high frequency fluctuations in $u_{ct}$. Also, we want to control for the effect of national business cycles on the fluctuations of $u_{ct}$. For these two reasons, our only interest is the frequencies within Band 2. As we stated above, this band consists of frequencies associated with cycles longer than 4 months but shorter than 18 months. We use Corbae and Ouliaris (2002)'s filter to remove all frequencies that lie in either Band 1 or Band 3. Corbae and Ouliaris's frequency domain filter also controls for any stochastic trend of unit root and involves no set up of parameter values.

After filtering $\{u_{ct}\}$, we obtain a new time series for each city $c$, denoted $u_{ct}^*$. Our spectral analysis is conducted on the frequency domain of $\{u_{ct}^*\}$.

## 4.2 Spectral Analysis

The spectral analysis reveals how cycles with different frequencies account for the fluctuations in a city's unemployment rate. A frequency of $\omega$ is associated with a cycle of length of $2\pi/\omega$. Let $s_y(\omega)$ be the power spectral density at $\omega$ of a time series $y$; $\int_0^{2\pi} s_y(\omega)d\omega$ is the total energy contained in fluctuations in $y$, denoted $G_y$. Thus, $\int_{\omega-\delta}^{\omega+\delta} s_y(f)df$ represents the portion of the energy that is attributed to frequencies that lie within the $\delta$-interval of frequency $\omega$. This reflects how much frequencies within that interval contribute to fluctuations in $y$.

We estimate the power spectrum density for each city. For city $c$ and a given $\delta_c$, we find a frequency $\omega$ whose $\delta_c$-interval contributes the most to the energy of $\{u^*_{ct}\}$. This frequency contributes more to the fluctuation in the city's unemployment rate than any of the other frequencies.

Formally, we define city $c$'s max-frequency as:

$$\omega_c^{max} = arg \max_{\delta_c \leq \omega \leq \pi - \delta_c} \int_{\omega - \delta_c}^{\omega + \delta} \theta(|\omega - f|) s_{u_c^*}(f) df,$$

where $\theta(\cdot)$ is a weight function. We let:

$$\theta(f) = \begin{cases} 0, & \text{if} \quad |f - w| > \delta_c; \\ \frac{0.8^{2|f-w|/\delta_c}}{\int_{\omega-\delta_c}^{\omega+\delta_c} 0.8^{2|f-w|/\delta_c} df}, & \text{if} \quad |f - \omega| \leq \delta_c. \end{cases}$$

This weight function has the property that the closer the frequency $f$ is to $\omega$, the larger the weight assigned to this frequency is.

Selecting an appropriate $\delta_c$ depends on how smooth the power spectral density curve of time series $\{u^*_{ct}\}$ is and what method is used to estimate the power spectral density. A smaller $\delta_c$ implies less robustness but more accuracy in calculating the max-frequency. After some experiments, we choose $\delta_c = .049$, which is 3% of the whole spectral domain $[0, \pi]$. Another frequency we are interested in is given by:

$$\omega_c^{mean} = \int_0^{\pi} \frac{s_{u_c^*}(f)}{G_{u_c^*}} f df.$$

The variable $\omega_c^{mean}$ is called the "mean-frequency" since it is a weighted average of frequencies over the frequency domain where the weight of each frequency is its (normalized) power spectral density. The higher the mean-frequency, the more contributions from high frequency cycles to unemployment fluctuations there are.

Table 3 is a summary of statistics of the frequency variables. The max-frequency and mean-frequency are .631 and .793, corresponding to 10.0 months and 7.9 months, respectively.

Table 3: Summary Statistics of Frequency Variables

| | mean | std dev | minimum | maximum |
|---|---|---|---|---|
| max-frequency | 0.631 | 0.235 | 0.368 | 1.520 |
| mean-frequency | 0.793 | 0.0985 | 0.616 | 1.030 |

### 4.3 Results from Summary Regressions

In order to understand the spectral analysis conducted in this section, we present an example comparing the $\{unempr_t\}$, $\{u_t^*\}$, and the power spectrum of $\{u_t^*\}$ of two cities. The first city, Monroe, Louisiana (PMSA code 5200), is relatively small and has an average labor force figure of 52,589. The other city is Los Angeles (PMSA code 4480), with an average labor force figure of 3,532,300.

The example is illustrated in Figure 3. The first row in Figure 3 depicts the unemployment rate $unempr_{ct}$ for Monroe and Los Angeles. The average unemployment rate in Monroe is 8.30%; in Los Angeles, the average unemployment rate is 7.69%. This is consistent with the first prediction of our model.

The detrended and filtered unemployment rates $u_{ct}^*$ are illustrated in the second row figures. We will come back to these figures in Section 5. In the third row, we draw the power spectrum of $u_{ct}^*$ for both cities. The max-frequency in Monroe is .44, corresponding to a cycle of $2\pi/.44 =$ 14.3 months. For Los Angeles, its max-frequency occurs at 1.08, which corresponds to a cycle of around 6 months. The larger city has shorter cycles than the smaller city, consistent with the second prediction of our model.

We use simple regressions of the variables of max-frequency and mean-frequency on the log of city size to summarize the relationships. The results are shown in Table 4. As predicted by the model in Section 2, both the max-frequency and mean-frequency are significant and positively correlated with city size. To assess the magnitude of the effect of city size, consider an increase of two standard deviations in the log of city size. The max-frequency increases by 0.1. If the initial max-frequency is 0.631, which equals the mean of the max frequency across cities, the corresponding unemployment cycle will be shortened by 1.4 months. As to

Table 4: Frequency Regression Results

| dependent variable | max-frequency | mean-frequency |
|---|---|---|
| constant | .707 | .833 |
| | (.035) | (.014) |
| log(city size) | 0.050 | .026 |
| | (.019) | (.0077) |
| $R^2$ | 0.050 | 0.076 |
| No. of obs. | 139 | 139 |

the mean-frequency, it increases by 0.052. If the initial mean frequency is 0.793, which equals the mean of the mean-frequency across cities, the corresponding unemployment cycle will be shortened by .6 months. In summary, the results in this section support the second prediction of our model: a larger city has a shorter unemployment cycle.

## 5 The Duration Analysis

In this section we carry out a different experiment: we investigate the duration of cyclical fluctuations in the unemployment rate city by city. Following Diebold and Rudebusch (1990), "duration" here refers to the length of each cycle, while a "cycle" is the time length between two consecutive turning points of an unemployment rate. We will define the turning points later in this section.

The duration analysis differs from the spectral analysis in two aspects. First, in duration analysis, identifying turning points of a cycle depends on the subjective rule we use. In the spectral analysis, a cycle is defined in the strict sense of periodicity. Thus, the results of the spectral analysis do not depend on the rule used to identify the turning points of a cycle. Second, the results from the spectral analysis are concerned with a whole cycle. Thus, it is impossible to discern different behaviors at different stages of a cycle. In contrast, the duration analysis reveals the relationship between city size and the length of both trough-to-peak cycles and peak-to-trough cycles.

### 5.1 Duration of Unemployment Cycles

We examine the duration of cyclical fluctuations of $\{u_{ct}^*\}$, for each city $c$, where $\{u_{ct}^*\}$ is the detrended and filtered unemployment rate defined in the previous section. A cycle of $\{u_{ct}^*\}$ is the time length between two consecutive turning points of $\{u_{ct}^*\}$. The following are some useful definitions.

- A *trough point* is the point where an upturn is about to start. Because we are considering the unemployment rate, an upturn in $\{u_{ct}^*\}$ signals a downturn in the economy.
- A *peak point* is the point with the highest value of $\{u_{ct}^*\}$ between two consecutive trough points.
- A *trough-to-trough duration* is the length between two consecutive trough points.
- A *trough-to-peak duration* is the length between a trough point and the first peak point right after it.
- A *peak-to-trough duration* is the length between a peak point and the first trough point right after it.

The key issue, then, is how to identify an upturn in $\{u_{ct}^*\}$. The classic criterion for identifying a downturn in business cycles is the "two consecutive declines" rule associated with GDP. Here, we apply a similar criterion (with a slight modification) to determine unemployment cycles. Specifically, an upturn is signaled either by two consecutive periods of growth in the unemployment rate or by three consecutive time periods where each has a higher unemployment rate than the preceding; moreover, there should be at least two periods of growth in the unemployment rate in these three time periods. The modification made here is to control for small noises in the time series.

Table 5: Summary Statistics of Duration Variables

| | mean | std dev | min | max |
|---|---|---|---|---|
| trough-to-trough | 7.01 | 1.01 | 5.27 | 10.9 |
| trough-to-peak | 3.4 | 0.43 | 2.56 | 4.95 |
| peak-to-trough | 3.6 | 0.71 | 2.57 | 6.89 |
| trough rate | -0.613 | 0.383 | -3.21 | -0.172 |
| peak rate | 0.61 | 0.375 | 0.188 | 2.76 |

According to the above criterion, time $t$ is a trough point of $\{u^*_{ct}\}$ if and only if:

$$\begin{cases} \left(u^*_{ct-2} > u^*_{ct}, u^*_{ct-1} > u^*_{ct}\right) \text{ and} \\ u^*_{ct+1} > u^*_{ct} \text{ and} \\ u^*_{ct+2} > u^*_{ct+1} \text{ or} \left(u^*_{ct+2} > u^*_{ct}, u^*_{ct+3} > u^*_{ct+2}\right). \end{cases} \tag{19}$$

## 5.2 City Size and Durations of Unemployment Cycles

We first identify each city's peak and trough points of unemployment cycles according to (19). Next we calculate each city's average trough-to-trough duration, trough-to-peak duration, peak-to-trough duration, peak unemployment rate and trough unemployment rate. Table 5 provides a summary of statistics of these variables. The average length of cycles, measured by the average trough-to-trough duration, is 7.01 months. Note, in Table 3, the mean-frequency is .793, corresponding to a length of 7.9 months. The difference between the two measures arises from the fact that different methods are used to measure the cycles.

Table 6 lists results from some simple regressions that summarize the relationships between the log of city size and the unemployment cycles. In the first column, the trough-to-trough duration, i.e., the length of a whole cycle, is significantly and negatively correlated with city size. In particular, an increase in two standard deviations of the log of city size will result in a decrease in the duration of unemployment cycles by .66 months. If we decompose the entire cycle into two parts, Table 6 shows that the trough-to-peak duration is significantly and

Table 6: Regression Results of Duration Analysis

| | trough-to-trough | trough-to-peak | peak-to-trough |
|---|---|---|---|
| constant | 6.49 | 3.19 | 3.30 |
| | (.145) | (.062) | (.103) |
| log(city size) | -0.308 | -0.110 | -0.198 |
| | (.078) | (.033) | (.055) |
| $R^2$ | 0.102 | 0.073 | 0.087 |
| No. of obs. | 129 | 129 | 129 |

Table 7: Regression Results of Peak Unemployment Rate

| | peak rate | trough rate |
|---|---|---|
| constant | .423 | -.434 |
| | (.055) | (.055) |
| log(city size) | -.121 | .116 |
| | (.029) | (.030) |
| $R^2$ | .101 | .115 |
| No. of obs. | 129 | 129 |

negatively correlated with city size, while the peak-to-trough duration is also significantly and negatively correlated with city size.

The test results are consistent with the thick market model presented in Section 2. According to the second prediction of our model, a larger city in general has a shorter trough-to-peak duration. Due to the thick market effect, a city's unemployed workers accumulate before the local labor market reaches a large enough size to have workers actively search for jobs. A larger city typically needs less time to reach that market size, which implies a shorter trough-to-peak duration.

It is worth pointing out that there is a significant negative correlation between the peak-to-trough duration and city size. Our model has not yet considered precisely how the matching between firms and workers proceeds after the minimum critical size $\bar{n}^*$ is reached. The explanation of this fact is an interesting topic for future research.

As to the relationship between city size and peak unemployment rates, it is clear from Table 7 that the peak unemployment rate is significantly and negatively correlated with city size. This result supports the third prediction of the model in Section 2. A lower peak unemployment rate in a larger city indicates a shallower recession in that city. In particular, an increase in two standard deviations in the log of city size drops the peak unemployment rates by .26 percentage points. Table 7 also shows a significant and positive correlation between the trough point unemployment rate and city size. This means that in general a larger city has a milder expansion than a smaller city. This interesting phenomenon merits further investigation. In summary, the difference between the peak unemployment rate and the trough unemployment rate is smaller in a larger city than in a smaller city. In particular, an increase of two standard deviations in the log of city size results in a decrease of the difference by .5 percentage points.

Figure 3 graphically shows the durations in unemployment cycles; the second row graphs in Figure 3 show how the peak unemployment rate is negatively related to city size. The city of Monroe, Louisiana has much a larger peak unemployment rate than that of the much larger city of Los Angeles. The average peak rate in Monroe is .801, while the average peak rate in Los Angeles is .451. Moreover, the average length of cycles is also longer in Monroe than those of Los Angeles.

## 6 Conclusion

This paper explores the relationship between city size and the pattern of unemployment rate fluctuations. We present a model of the local labor market in which in a thicker market, when more workers are looking for jobs and more job openings are available, the matching probability between jobs and workers is better. Workers incur search cost if they actively search for for jobs. A higher matching probability makes searching for jobs more desirable. Unemployed workers accumulate in a local market until the market reaches a critical size such that the expected wage is higher than the search cost. Since a given shock produces more unemployed workers in a larger city during a given time period, it takes less time for a larger city to reach the critical size described above. As a consequence, the model predicts: (1) Unemployment rates are lower in larger cities. (2) The length of unemployment cycles decreases as city size increases. (3) The peak unemployment rate is negatively correlated with city size.

Our empirical analysis utilizes data that covers 295 PMSAs in the U.S. over the years 1981–1997. After controlling for the effects of industry composition and risk diversification, we find that city size has a significantly negative effect on the mean unemployment rate. In particular, if city size increases by two standard deviations, the unemployment rate will be lowered by roughly a half percentage point. We also find that larger cities have shorter unemployment cycles. In particular, the unemployment cycle will be shortened by roughly one month if city size increases by two standard deviations. Finally, we find shallower recessions for larger cities. All these empirical results are consistent with the predictions of the model.

## References

Abraham, K. G. and Shimer, R.: 2001, Changes in unemployment duration and labor force attachment, NBER working paper, #8513.

Corbae, D. and Ouliaris, S.: 2002, Extracting cycles from nonstationary data, Working Paper, Department of Economics, University of Texas at Austin.

Corbae, D., Ouliaris, S. and Phillips, P.: 2002, Band spectral regression with trending data, *Econometrica* **70**(3), 1067–1109.

Diamond, P.: 1982, Aggregate demand management in search equilibrium, *Journal of Political Economy* **90**(5), 881–894.

Diebold, F. X. and Rudebusch, G. D.: 1990, A nonparametric investigation of duration dependence in the American business cycle, *Journal of Political Economy* **98**(3), 596–616.

Gan, L. and Li, Q.: 2002, Efficiency of thin vs. thick markets, Working Paper, Department of Economics, University of Texas Austin.

Helsley, R. and Strange, W.: 1990, Matching and agglomeration economies in a system of cities, *Regional Science and Urban Economics* **20**, 189–212.

Helsley, R. and Strange, W.: 1991, Agglomeration economies and urban capital markets, *Journal of Urban Economics* **29**, 96–112.

Henderson, V. J.: 1986, The efficiency of resource usage and city size, *Journal of Urban Economics* **19**, 47–70.

Henderson, V. J.: 1988, *Urban Economics*, New York: Oxford University Press.

Howitt, P. and McAfee, P. R.: 1987, Costly search and recruiting, *International Economic Review* **28**(1), 89–107.

Mills, E. J. and Hamilton, B. W.: 1984, *Urban Economics*, Glenview, IL: Scott, Foresman and Company.

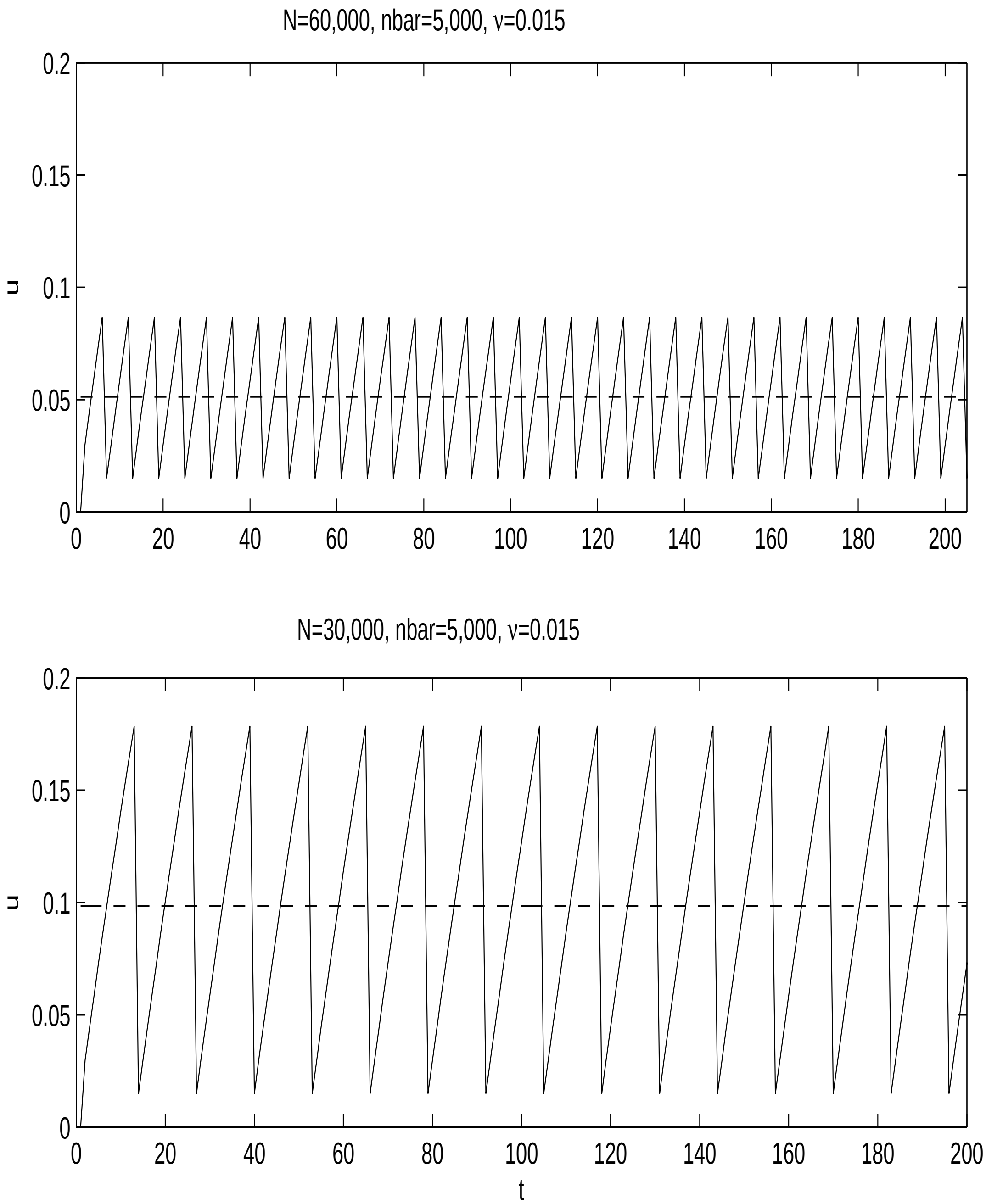
Neumann, G. and Topel, R.: 1991, Employment risk, diversification, and unemployment, *Quarterly Journal of Economics* **106**(4), 1341–65.

Simon, C.: 1988, Frictional unemployment and the role of industrial diversity, *Quarterly Journal of Economics* **103**(4), 715–28.

Wilson, P.: 1988, Wage variation resulting from staggered work hours, *Journal of Urban Economics* **24**, 9–12.

Zhang, Q.: 2002, Microfoundations of local business cycles, Working Paper, Department of Economics, University of Texas at Austin.

Figure 1: Illustration of the Thick Market Effect on Unemployment Fluctuations



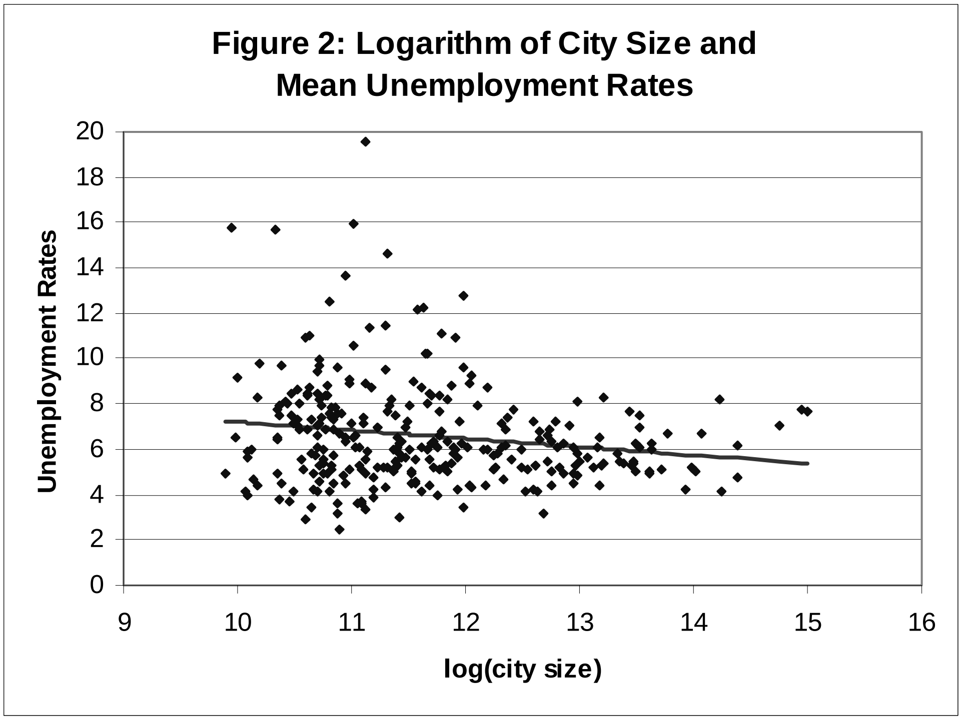


Figure 2: Logarithm of City Size and
Mean Unemployment Rates
Unemployment Rates
20
18
16
14
12
10
8
6
4
2
0
9
10
11
12
13
14
15
16
log(city size)

Figure 3: Patterns of Unemployment Rates in Monroe, LA, and Los Angeles, CA