ECONOMICS AND THE CITY

Jan Eeckhout†

†Barcelona GSE-UPF

Bojos per l’Economia
31 January, 2015
Cities

Labor markets

Local mobility

population dynamics

Zipf's Law

Gibrat's Law

wages

housing prices

productivity differences

geographical: mountains and waterways

agglomeration externalities

Alfred Marshall

Urban Wage Premium

skills

Sorting

Policy

Income Taxes

Optimal Spatial Taxation
Cities

Labor markets
Cities

Local Labor markets
Cities

Local Labor markets

mobility

CITIES
Local Labor markets
CITIES

Local Labor markets
Gibrat’s Law Zipf’s Law

mobility
population dynamics

mobsity
population dynamics

wages
housing prices
productivity differences
geographical: mountains and waterways
agglomeration externalities

Alfred Marshall
Urban Wage Premium

skills
Sorting
Policy
Income Taxes
Optimal Spatial Taxation
CITIES

Local Labor markets

mobility

Gibrat’s Law

Zipf’s Law

housing prices

wages

mobility

population dynamics

productivity differences

generally geographical: mountains and waterways

agglomeration externalities

Alfred Marshall

Urban Wage Premium

skills

Sorting

Policy

Income Taxes

Optimal Spatial Taxation
CITIES

Local Labor markets

wages

housing prices

productivity differences

mobility

Gibrat’s Law

Zipf’s Law

mobility

population dynamics

geographical: mountains and waterways

agglomeration externalities

Alfred Marshall

Urban Wage Premium

skills

Sorting

Policy

Income Taxes

Optimal Spatial Taxation
Cities

Local Labor markets

Gibrat’s Law  Zipf’s Law

housing prices
wages

productivity differences

geographical: mountains and waterways

mobility

mobility

population dynamics

Optimal Spatial Taxation

Sorting

Policy

Income Taxes

Urban Wage Premium

Alfred Marshall

skills
Cities
Labor markets
Local mobility
population dynamics
Zipf’s Law
Gibrat’s Law
wages
housing prices
productivity differences
geographical: mountains and waterways
agglomeration externalities
Alfred Marshall
skills
Policy
Income Taxes
Sorting
wages
Urban Wage Premium
mobility
Local Labor markets
Sorting
Income Taxes
Policy
agglomeration externalities
productivity differences
geographical: mountains and waterways
populations dynamics
Zipf’s Law
Gibrat’s Law
Cities
Labor markets
Local mobility
population dynamics
Zipf's Law
Gibrat's Law
wages
housing prices
productivity differences
geographical: mountains and waterways
agglomeration externalities
Alfred Marshall
Urban Wage Premium
Optimal Spatial Taxation
Policy
Sorting
Income Taxes
wages
mobility
Local Labor markets
Sorting
Income Taxes
skills
CITIES
Policy
Skills
Sorting
Income Taxes
wages
mobility
Local Labor markets
Gibrat's Law
Zipf's Law
Optimal Spatial Taxation
Productivity differences
Geographical: mountains and waterways
Agglomeration externalities
Alfred Marshall
1. Zipf’s and Gibrat’s law: where does it come from?
Three Questions

1. Zipf’s and Gibrat’s law: where does it come from?
   - A Surprising Regularity and a puzzle
   - Economic forces
Three Questions

1. Zipf’s and Gibrat’s law: where does it come from?
   • A Surprising Regularity and a puzzle
   • Economic forces
2. Is there Spatial Sorting?
Three Questions

1. Zipf’s and Gibrat’s law: where does it come from?
   • A Surprising Regularity and a puzzle
   • Economic forces

2. Is there Spatial Sorting?
   • Who works in big cities?
   • Technological determinants
THREE QUESTIONS

1. Zipf’s and Gibrat’s law: where does it come from?
   - A Surprising Regularity and a puzzle
   - Economic forces
2. Is there Spatial Sorting?
   - Who works in big cities?
   - Technological determinants
3. Does Federal Income Taxation affect local labor markets?
Three Questions

1. Zipf’s and Gibrat’s law: where does it come from?
   - A Surprising Regularity and a puzzle
   - Economic forces

2. Is there Spatial Sorting?
   - Who works in big cities?
   - Technological determinants

3. Does Federal Income Taxation affect local labor markets?
   - Effect on location decisions
   - Optimal taxation policy
Introduction

- Why are there cities of sizes? Why are there cities?
  - Geographical determinants? Rivers, weather,...
  - Consumer demand: amenities from size? Opera,...
  - Labor markets?
- What are the technological determinants of productivity across different size cities?
Introduction

- Why are there cities of sizes? Why are there cities?
  - Geographical determinants? Rivers, weather,...
  - Consumer demand: amenities from size? Opera,...
  - Labor markets?
- What are the technological determinants of productivity across different size cities?
- Address two puzzles + policy implications:
  1. Proportionate growth and Zipf’s law
  2. Urban Wage Premium
  3. Taxation
- Exploit the relation: wages – population – housing prices
I  Zipf’s and Gibrat’s law
II  Spatial Sorting
III  Taxation
I. Population and Labor Market Dynamics

Zipf’s law

**Figure I**

Log Size versus Log Rank of the 135 largest U.S. Metropolitan Areas in 1991

Source: Statistical Abstract of the United States [1993].

\[(1) \quad \ln \text{Rank} = 10.53 - 1.005 \ln \text{Size}, \quad (.010)\]
Zipf’s Law

- The largest city is $N$ times larger than the $N$-th city

$$S \approx \frac{e^a}{\text{Rank}} \quad (a = 10.53)$$

- First observed by Zipf (1949)
- Early systematic pattern: Le Maître (1648), Auerbach (1913)
- Robust across time and space
Zipf’s Law

- The largest city is $N$ times larger than the $N$-th city

\[ S \approx \frac{e^a}{\text{Rank}} \quad (a = 10.53) \]

- First observed by Zipf (1949)
- Early systematic pattern: Le Maître (1648), Auerbach (1913)
- Robust across time and space
- Remarkably systematic relationship

$\Rightarrow$ Krugman (1995): “We have to say that the rank-size rule is a major embarrassment for economic theory: one of the strongest statistical relationships we know, lacking any clear basis in theory.”
ZIPF’S LAW

Table 2—Ten Largest Metropolitan Areas in the United States

<table>
<thead>
<tr>
<th>Rank</th>
<th>MA</th>
<th>Population</th>
<th>$S_{NY}/S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>New York-Northern New Jersey-Long Island, NY-NJ-CT-PA</td>
<td>21,199,865</td>
<td>1.000</td>
</tr>
<tr>
<td>2</td>
<td>Los Angeles-Riverside-Orange County, CA</td>
<td>16,373,645</td>
<td>1.295</td>
</tr>
<tr>
<td>3</td>
<td>Chicago-Gary-Kenosha, IL-IN-WI</td>
<td>9,157,540</td>
<td>2.315</td>
</tr>
<tr>
<td>4</td>
<td>Washington-Baltimore, DC-MD-VA-WV</td>
<td>7,608,070</td>
<td>2.787</td>
</tr>
<tr>
<td>5</td>
<td>San Francisco-Oakland-San Jose, CA</td>
<td>7,039,362</td>
<td>3.012</td>
</tr>
<tr>
<td>6</td>
<td>Philadelphia-Wilmington-Atlantic City, PA-NJ-DE-MD</td>
<td>6,188,463</td>
<td>3.426</td>
</tr>
<tr>
<td>7</td>
<td>Boston-Worcester-Lawrence, MA-NH-ME-CT</td>
<td>5,819,100</td>
<td>3.643</td>
</tr>
<tr>
<td>8</td>
<td>Detroit-Ann Arbor-Flint, MI</td>
<td>5,456,428</td>
<td>3.885</td>
</tr>
<tr>
<td>9</td>
<td>Dallas-Fort Worth, TX</td>
<td>5,221,801</td>
<td>4.060</td>
</tr>
<tr>
<td>10</td>
<td>Houston-Galveston-Brazoria, TX</td>
<td>4,669,571</td>
<td>4.540</td>
</tr>
</tbody>
</table>

Note: $S_{NY}/S$ denotes the ratio of population size relative to New York.
Source: Census Bureau, 2000.
Two open questions:

1. Why Pareto distribution?
   - Pareto vs. other distributions?
   - Why so robust?

2. What are the economic forces behind this?
**ZIPF’S LAW**

• Zipf’s law: size distribution is Pareto with scale coefficient 1

• Pareto distribution ($\forall S \geq S$):

\[
p(S) = \frac{aS^a}{S^{a+1}}
\]

\[
P(S) = 1 - \left(\frac{S}{S_S}\right)^a
\]

• If we denote rank by $r$, then (where $\underline{N}$ is $\#$ cities above cutoff):

\[
r = \underline{N}(1 - P(S)) = \underline{N} \left(\frac{S}{S_S}\right)^a
\]

and therefore

\[
\ln r = K - a \ln S
\]

(\text{where } K = \ln \underline{N} + a \ln S).
Zipf's law

The only remaining issue to resolve is how it is possible that Zipf's law is repeatedly confirmed in the literature, while the underlying distribution is lognormal. The Pareto distribution is very different from the lognormal, so it is obvious that if the true distribution is lognormal, the entire distribution can never be fit to a Pareto distribution at the same time. Consider Figure 1 with a plot of the density function of the lognormal and that of the Pareto distribution (both on a ln scale); observe that the lognormal on a log scale is the normal density function. The density of the Pareto distribution is downward sloping, whereas the lognormal density is initially increasing and then decreasing (given symmetry, half the observations are in the increasing part). If the underlying distribution is lognormal, then goodness of fit tests will categorically reject the Pareto distribution. Still, when regressing log rank on log size for the entire distribution, the coefficient comes out significant. Estimating a linear coefficient when the underlying empirical distribution is not Pareto (i.e., the relation is nonlinear) can obviously produce a significant estimate. This regression test merely confirms that there is a relation between size and rank, but it does not provide a test for the linearity of this relation. As such, testing the significance of the linear coefficient is not the equivalent of a goodness-of-fit test for the Pareto distribution.

More important though is that until now the literature considered the truncated distribution (typically, the truncation point is at ln size equal to 12 on the horizontal axis, i.e., for only 135 cities). At the very upper tail of the distribution, there is no dramatic difference between the density function of the lognormal and the Pareto. Now both the truncated lognormal and the Pareto density are downward sloping and similar (the Pareto is slightly more convex). As a result, both the Pareto and the truncated lognormal trace the data relatively closely.
A SECOND REGULARITY
PROPORTIONATE GROWTH

- Cities grow at different rates
- Growth is stochastic
- But: the average growth rate is independent of size
A SECOND REGULARITY
PROPORTIONATE GROWTH

20
Unfortunately, these restrictions do not allow for the possibility of augmenting the dataset to include populations that are currently not covered.

21
It should be noted that the current dataset of all cities has already been augmented to form the largest possible dataset that is feasible, with the inclusion of the census-defined CDPs. This increases the number of cities by 31 percent, from 19,361 to 25,359.

22
The fact that part of the population is not covered is potentially a cause for concern, because rather than capturing deep patterns of populations and population dynamics, we may merely be describing the idiosyncrasy of the jurisdictional formation in the United States. The population that is not covered may be distributed in a completely different way from the lognormal distribution. And since we cannot assign that population to any geographic area comparable to a city, there is no hope of knowing how the remainder is distributed. The lognormality seems to be a strong regularity, however, from whichever perspective population dynamics is considered. First, while we have no way of showing that the distribution of MAs is lognormal given the truncation by definition, we show below that even for MAs, changes in the truncation point produce changes in the estimated Zipf coefficient that are consistent with the fact that the underlying upper tail is derived from the lognormal. Second, the size distribution of CPDs is pretty close to the entire distribution of cities and hence the lognormal. And finally, in the Appendix we show the results of further analysis using additional data that are available from the Census. We plot the size distribution of counties, which covers the entire U.S. population (see Figure A-1 and Table A-1 in Appendix A for the ten largest counties). While it is hardly convincing to make a case for counties as the relevant economic unit, it is surprising that even the size distribution of counties is close to the lognormal. Looking at population dynamics from the perspective of different economic units and including as large a fraction as possible of the U.S. population, there is a strong pattern that is consistent with lognormality.

C.
PROPORTIONATE CITY GROWTH

For the cities in the upper tail of the size distribution, population growth has repeatedly been shown to satisfy constant proportionate growth.

These findings can be extended beyond those for the upper tail of the distribution. We therefore use the data on population size for places in the United States from both the 1990 and 2000 Censuses. Unfortunately, 1990 Census data do not include the CDPs. As a result, the sample size is significantly smaller (19,361 instead of 25,359). Figure 4 shows the scatter plot of growth against city size (on ln scales).

While it is hardly convincing to make a case for counties as the relevant economic unit, it is surprising that even the size distribution of counties is close to the lognormal.
A Second Regularity
Proportionate Growth

Support that growth is independent of size. In what follows, the dependence relation of growth on size is analyzed in greater detail. We perform both nonparametric and parametric regressions of growth on size.

First, we perform a nonparametric regression of growth on size. The standard parametric regressions as performed below provide us only with an aggregate relationship between growth and size, which is constrained to hold over the entire support of the distribution of city sizes. In contrast, the nonparametric estimate allows growth to vary with size over the distribution. The regression relationship we model is therefore

\[
g_i = \frac{\gamma}{H_1} \frac{\gamma}{H_2} S_i \frac{\gamma}{H_3} \frac{\gamma}{H_4} \frac{\gamma}{H_5} i \text{ for all } i = 1, \ldots, 19361.
\]

The objective is to provide an approximation of the unknown relationship between growth and size using smoothing, without making parametric assumptions about the functional form of \( m \). Before estimating \( m \), we report the distribution of growth rates for each decile of the size distribution. Following Ioannides and Overman (2003), we use the normalized growth rate (the difference between the growth rate and the sample mean divided by the standard deviation). In Figure 5, the stochastic kernel density is plotted for each of the 10 deciles. Fixing a particular decile in the distribution, we can observe the distribution of growth rates within that decile. Figure 6 reports the contour plot of the same stochastic kernel, i.e., the vertical projection of the density function. Both figures illustrate that the distribution of growth rates is strikingly stable over different deciles. The best illustration of the size independence is the fact that the contour lines are parallel. The distribution is slightly skewed (the mode is just below zero), and the mode appears fairly constant over different deciles. The same is true for the variance. While the variance of the lowest decile seems to be somewhat higher (the contour lines fan out somewhat), there seems to be little change in the spread of the distribution for higher deciles.

We now proceed to estimate the regression relationship

\[
g_i = \frac{\gamma}{H_1} m(S_i) \frac{\gamma}{H_2} \frac{\gamma}{H_3} \frac{\gamma}{H_4} \frac{\gamma}{H_5} i \text{ for all } i = 1, \ldots, 19361,
\]

where \( g_i \) is the normalized growth rate, i.e., the

This section on the nonparametric analysis follows closely the analysis in Ioannides and Overman (2003). We derive a sequence of results for our dataset of all cities similar to theirs, obtained for a time-series dataset on the largest MAs.

Each stochastic kernel is calculated using the bandwidth derived with the automatic method corresponding to the Gaussian distribution (see Bernard W. Silverman, 1986).
A SECOND REGULARITY
PROPORTIONATE GROWTH

The difference between growth and the sample mean divided by the sample standard deviation, and \( S_i \) is the log of the population size of a city. We will approximate the true relationship by the regression curve \( m(s) \) for all \( s \) in the support of \( S_i \). The estimate of \( m(s) \) will be denoted \( \hat{m}(s) \) and is a local average around the point \( s \). This local average smooths the value around \( s \), and the smoothing is done using a kernel, i.e., a continuous weight function symmetric around \( s \).

The kernel \( K \) used in the remainder of the paper will be an Epanechnikov kernel. The bandwidth \( h \) determines the scale of the smoothing, and \( K_h \) denotes the dependence of \( K \) on the bandwidth \( h \). With the kernel weights, we calculate the estimate of \( m \) using the Nadaraya-Watson method, where

\[
\hat{m}(s) = \frac{\sum_{i=1}^{n} K_h(s - s_i) g_i}{\sum_{i=1}^{n} K_h(s - s_i)}
\]

In Figure 7 there is a plot of \( \hat{m}(s) \) calculated for a bandwidth of \( h = 0.5 \) (see Silverman, 1986). The Figure also shows the bootstrapped 95-percent confidence bands (calculated from 500 random samples with replacement). In line with the earlier results, the nonparametric estimate of the conditional mean is stable across different population sizes, except for the very bottom of the distribution. The estimate seems to exhibit some slightly inverted U-shape, with somewhat higher growth rates in the middle range of population sizes and lower growth at the ends. If the underlying relation between growth and size is constant, then the estimate will lie in the 95-percent confidence bands. This seems to suggest that, except for some values near the lower boundary, we cannot reject that growth is independent of size. Observe that because the kernel is a fixed function and boundary observations have support only on one side of the kernel, the kernel estimates near the boundaries must be read with caution.

In Table B-1 in Appendix B, some further descriptive statistics are reported for growth rates over the entire support of the distribution. Consistent with the kernel estimates, average growth rates seem to be constant, except at the very bottom of the distribution. We also calculate the standard deviation and the Interquartile Range (IQR) of the growth rate. The IQR is defined as the difference between the seventy-fifth and twenty-fifth percentiles (\( Q_3 - Q_1 \)). This provides an indication of the variation in growth rates. For the largest 100 cities, growth rates vary less, whereas the smallest 100 cities exhibit higher variation in growth rates. The results below have been replicated using the Gaussian kernel and reveal no differences with those using the Epanechnikov kernel.

See Wolfgang H"{o}rdle (1990). At the bottom of the distribution there is also more variation in growth rates (see IQR calculations below). Because the confidence bands impose a requirement over the entire domain of the size distribution, the width of the bands is likely to be affected by the variation at the bottom.
A Second Regularity
Proportionate Growth

The difference between growth and the sample mean divided by the sample standard deviation, and \( S_i \) is the log of the population size of a city. We will approximate the true relationship by the regression curve \( m(s) \) for all \( s \) in the support of \( S_i \). The estimate of \( m(s) \) will be denoted \( \hat{m}(s) \) and is a local average around the point \( s \). This local average smooths the value around \( s \), and the smoothing is done using a kernel, i.e., a continuous weight function symmetric around \( s \).

The kernel \( K \) used in the remainder of the paper will be an Epanechnikov kernel. The bandwidth \( h \) determines the scale of the smoothing, and \( K_h \) denotes the dependence of \( K \) on the bandwidth \( h \).

With the kernel weights, we calculate the estimate of \( m \) using the Nadaraya-Watson method, where

$$
\hat{m}(s) = \frac{1}{n} \sum_{i=1}^{n} K_h(s - S_i) g(i)
$$

In Figure 7 there is a plot of \( \hat{m}(s) \) calculated for a bandwidth of \( h = 0.5 \) (see Silverman, 1986). The Figure also shows the bootstrapped 95-percent confidence bands (calculated from 500 random samples with replacement). In line with the earlier results, the nonparametric estimate of the conditional mean is stable across different population sizes, except for the very bottom of the distribution. The estimate seems to exhibit some slightly inverted U-shape, with somewhat higher growth rates in the middle range of population sizes and lower growth at the ends. If the underlying relation between growth and size is constant, then the estimate will lie in the 95-percent confidence bands. This seems to suggest that, except for some values near the lower boundary, we cannot reject that growth is independent of size. Observe that because the kernel is a fixed function and boundary observations have support only on one side of the kernel, the kernel estimates near the boundaries must be read with caution.

In Table B-1 in Appendix B, some further descriptive statistics are reported for growth rates over the entire support of the distribution. Consistent with the kernel estimates, average growth rates seem to be constant, except at the very bottom of the distribution. We also calculate the standard deviation and the Interquartile Range (IQR) of the growth rate. The IQR is defined as the difference between the seventy-fifth and twenty-fifth percentiles (\( Q_3 - Q_1 \)). This provides an indication of the variation in growth rates. For the largest 100 cities, growth rates vary less, whereas the smallest 100 cities exhibit higher variation in growth rates. The results below have been replicated using the Gaussian kernel and reveal no differences with those using the Epanechnikov kernel.

26 See Wolfgang Härdle (1990).

27 At the bottom of the distribution there is also more variation in growth rates (see IQR calculations below). Because the confidence bands impose a requirement over the entire domain of the size distribution, the width of the bands is likely to be affected by the variation at the bottom.
**Proportionate Growth**

Parametric growth regressions:

\[
\frac{S_{00}}{S_{90}} = 1.102 - 3.75E(-08) \frac{S_{90} + S_{00}}{2} \\
(0.005) (7E(-08))
\]

\[
\frac{S_{00}}{S_{90}} = 1.103 + 2.3E(-09)S_{90} \\
(0.005) (7.3E(-08))
\]
A Puzzle

- How can we reconcile
  1. Zipf’s law, and
  2. proportionate growth?
A Puzzle

• How can we reconcile
  1. Zipf’s law, and
  2. proportionate growth?

• Reason: Gibrat’s Law: proportionate growth
  $\Rightarrow$ log-normal distribution of city sizes, not Pareto

• Proportionate growth

$$S_{i,t} = (1 + \varepsilon_{i,t})S_{i,t-1}$$

$$\sum_{t=1}^{T} \frac{S_{i,t} - S_{i,t-1}}{S_{i,t-1}} = \sum_{t=1}^{T} \varepsilon_{i,t}$$

$$\sum_{t=1}^{T} \frac{S_{i,t} - S_{i,t-1}}{S_{i,t-1}} \approx \int_{S_{i,0}}^{S_{i,T}} \frac{dS_i}{S_i} = \ln S_{i,t} - \ln S_{i,0}$$
A Puzzle

• Between any two periods $t$:

$$\ln S_{i,t} = \ln S_{i,t-1} + \varepsilon_{i,t}$$

and therefore:

$$\ln S_{i,T} = \ln S_{i,0} + \varepsilon_{i,1} + \cdots + +\varepsilon_{i,T}.$$  

• From the central limit theorem, $\ln S_{i,T}$ is asymptotically normal, and therefore $S_{i,T}$ is asymptotically log-normal (Gibrat 1931)

$\Rightarrow$ Proportionate growth $\Rightarrow$ lognormal distribution (not Pareto)
A Puzzle

Information concerning new technologies and products spills over faster in markets with high degrees of local interaction, like those of large cities. Simultaneously, workers in larger cities also impose negative externalities on each other because commuting times are longer. The economy differs from the one in Lucas and Rossi-Hansberg (2002) because of the explicit mobility between cities, rather than within cities. The aim is to capture the notion of competition between geographic locations, i.e., perfectly mobile citizens making location decisions between different cities. Local externalities within cities regulate the mobility of citizens between different cities (i.e., there are no externalities between cities). It is shown that the local externality model economy predicts behavior that is consistent with the empirical city growth process.

The only remaining issue to resolve is how it is possible that Zipf's law is repeatedly confirmed in the literature, while the underlying distribution is lognormal. The Pareto distribution is very different from the lognormal, so it is obvious that if the true distribution is lognormal, the entire distribution can never be fit to a Pareto distribution at the same time. Consider Figure 1 with a plot of the density function of the lognormal and that of the Pareto distribution (both on a ln scale); observe that the lognormal on a log scale is the normal density function. The density of the Pareto distribution is downward sloping, whereas the lognormal density is initially increasing and then decreasing (given symmetry, half the observations are in the increasing part). If the underlying distribution is lognormal, then goodness of fit tests will categorically reject the Pareto distribution. Still, when regressing log rank on log size for the entire distribution, the coefficient comes out significant. Estimating a linear coefficient when the underlying empirical distribution is not Pareto (i.e., the relation is nonlinear) can obviously produce a significant estimate. This regression test merely confirms that there is a relation between size and rank, but it does not provide a test for the linearity of this relation. As such, testing the significance of the linear coefficient is not the equivalent of a goodness-of-fit test for the Pareto distribution.

More important though is that until now the literature considered the truncated distribution (typically, the truncation point is at ln size equal to 12 on the horizontal axis, i.e., for only 135 cities). At the very upper tail of the distribution, there is no dramatic difference between the density function of the lognormal and the Pareto. Now both the truncated lognormal and the Pareto density are downward sloping and similar (the Pareto is slightly more convex). As a result, both the Pareto and the truncated lognormal trace the data relatively closely. The problem is...
Reconciling Evidence

- Gabaix (1999): a process with entry and exit at high truncation
- The fit of the Pareto tail (Zipf’s law) is for 135 cities only
  \[\Rightarrow\] something going on outside tail
  \[\Rightarrow\] need to consider entire distribution, not just the truncation
EVIDENCE
ZIPF’S LAW FOR (ALL) MSA’S?

The estimated coefficient on the Pareto distribution is clearly sensitive to the choice of the truncation point. Moreover, the dependence of the estimated coefficient of the Pareto distribution on the truncation point, other explanations for the size distribution of cities. In what follows, it is shown that a theoretical justification exists for the size distribution of cities. In particular, ln(1 + x) is clearly increasing for an increasing truncation point; see the review on MAs by Gabaix and Ioannides (2003).

As before, let the lognormal density function respond to changes in the Pareto coefficient and the truncation point. Other explanations for the size distribution of cities have been suggested in the literature. In particular, ln(1 + x) is inversely proportional to rank (Zipf’s law) is not required. The plot of the hazard rate for the corresponding lognormal distribution with sample size of the sample population increases. This establishes a decrease in the truncated sample population.

Because an increase in the truncation size implies a decrease in the truncated sample population, the estimate will be decreasing as the sample size of the sample population increases. This establishes a decrease in the truncated sample population. Moreover, the dependence of the estimated coefficient of the Pareto distribution on the truncation point is clearly sensitive to the choice of the truncation point.
Evidence
Places

- By definition, MSA is truncated (at least one city with population $> 50,000$)
- Use a different definition: incorporated places
  - Largest: five boroughs of NYC
  - But not New Jersey, Connecticut,...
  - Based on the legal definition (mayor,...)
  - Some are extremely small (zero population!)
  - 25,359 places; median size = 1,338
  - Only 73% of population
### Evidence Places

Table 1—Ten Largest Cities in the United States

<table>
<thead>
<tr>
<th>Rank</th>
<th>City</th>
<th>Population S</th>
<th>$S_{NY}/S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>New York, NY</td>
<td>8,008,278</td>
<td>1.000</td>
</tr>
<tr>
<td>2</td>
<td>Los Angeles, CA</td>
<td>3,694,820</td>
<td>2.167</td>
</tr>
<tr>
<td>3</td>
<td>Chicago, IL</td>
<td>2,896,016</td>
<td>2.753</td>
</tr>
<tr>
<td>4</td>
<td>Houston, TX</td>
<td>1,953,631</td>
<td>4.099</td>
</tr>
<tr>
<td>5</td>
<td>Philadelphia, PA</td>
<td>1,517,550</td>
<td>5.277</td>
</tr>
<tr>
<td>6</td>
<td>Phoenix, AZ</td>
<td>1,321,045</td>
<td>6.062</td>
</tr>
<tr>
<td>7</td>
<td>San Diego, CA</td>
<td>1,223,400</td>
<td>6.546</td>
</tr>
<tr>
<td>8</td>
<td>Dallas, TX</td>
<td>1,188,580</td>
<td>6.738</td>
</tr>
<tr>
<td>9</td>
<td>San Antonio, TX</td>
<td>1,144,646</td>
<td>6.996</td>
</tr>
<tr>
<td>10</td>
<td>Detroit, MI</td>
<td>951,270</td>
<td>8.419</td>
</tr>
</tbody>
</table>

*Note:* $S_{NY}/S$ denotes the ratio of population size relative to New York.

Evidence All Cities

Cities and MAs represent different notions about the corresponding theory of an economic unit. And depending on the definition, we are studying different objects and therefore different distributions. As is the case with comparisons of countries, we do not have a perfect justification for using a particular unit of account when comparing cities. In our theory below, we consider local externalities that do not affect agents outside the economic unit as the defining characteristic of a city. In reality of course, no externality is purely local. One may therefore want to interpret this assumption as a matter of the extent to which externalities do or do not affect agents outside a given city. The danger is that the partition into economic units is either too fine or, at the other extreme, too coarse. The externalities for some agents in one part of a given economic unit (say those living in New Haven) may not have an impact on those living in different parts of the same unit (say Princeton). Moreover, different research objectives may call for the use of different units of account. For example, if one is interested in analyzing the economic impact of airports, the MA seems a natural unit of account, while cities may be more appropriate when studying schools, public transportation, or waste collection. In past research, both MAs and cities have proven to be useful and relevant economic units, and both have been studied extensively. In this paper, cities are chosen for several reasons. In addition to the fact that cities are a natural economic unit for studying the local externalities that are modeled in Section III, there is a practical reason: the availability of data. We want to use data that cover the entire range of populations, in particular the smaller ones. Because MAs are defined by the Census Bureau only for large populations (MAs must include “at least one city with 50,000 or

Figure 2. Empirical and Theoretical Density Functions

Figure 3. Empirical and Theoretical Cumulative Density Functions

Table 2—Ten Largest Metropolitan Areas in the United States

<table>
<thead>
<tr>
<th>Rank</th>
<th>MA Population</th>
<th>Population Size Relative to New York</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Los Angeles-Riverside-Orange County, CA</td>
<td>16,373,645</td>
<td>1.295</td>
</tr>
<tr>
<td>3</td>
<td>Chicago-Gary-Kenosha, IL-IN-WI</td>
<td>9,157,540</td>
<td>2.315</td>
</tr>
<tr>
<td>4</td>
<td>Washington-Baltimore, DC-MD-VA-WV</td>
<td>7,608,070</td>
<td>2.787</td>
</tr>
<tr>
<td>5</td>
<td>San Francisco-Oakland-San Jose, CA</td>
<td>7,039,362</td>
<td>3.012</td>
</tr>
<tr>
<td>6</td>
<td>Philadelphia-Wilmington-Atlantic City, PA-NJ-DE-MD</td>
<td>6,188,463</td>
<td>3.426</td>
</tr>
<tr>
<td>7</td>
<td>Boston-Worcester-Lawrence, MA-NH-ME-CT</td>
<td>5,819,100</td>
<td>3.643</td>
</tr>
<tr>
<td>8</td>
<td>Detroit-Ann Arbor-Flint, MI</td>
<td>5,456,428</td>
<td>3.885</td>
</tr>
<tr>
<td>9</td>
<td>Dallas-Fort Worth, TX</td>
<td>5,221,801</td>
<td>4.060</td>
</tr>
<tr>
<td>10</td>
<td>Houston-Galveston-Brazoria, TX</td>
<td>4,669,571</td>
<td>4.540</td>
</tr>
</tbody>
</table>

Note: NY/S denotes the ratio of population size relative to New York.
Evidence
All Cities

Cities and MAs represent different notions about the corresponding theory of an economic unit. And depending on the definition, we are studying different objects and therefore different distributions. As is the case with comparisons of countries, we do not have a perfect justification for using a particular unit of account when comparing cities. In our theory below, we consider local externalities that do not affect agents outside the economic unit as the defining characteristic of a city. In reality of course, no externality is purely local. One may therefore want to interpret this assumption as a matter of the extent to which externalities do or do not affect agents outside a given city. The danger is that the partition into economic units is either too fine or, at the other extreme, too coarse. The externalities for some agents in one part of a given economic unit (say those living in New Haven) may not have an impact on those living in different parts of the same unit (say Princeton). Moreover, different research objectives may call for the use of different units of account. For example, if one is interested in analyzing the economic impact of airports, the MA seems a natural unit of account, while cities may be more appropriate when studying schools, public transportation, or waste collection. In past research, both MAs and cities have proven to be useful and relevant economic units, and both have been studied extensively. In this paper, cities are chosen for several reasons. In addition to the fact that cities are a natural economic unit for studying the local externalities that are modeled in Section III, there is a practical reason: the availability of data. We want to use data that cover the entire range of the populations, in particular the smaller ones. Because MAs are defined by the Census Bureau only for large populations (MAs must include “at least one city with 50,000 or

---

Table 2—Ten Largest Metropolitan Areas in the United States

<table>
<thead>
<tr>
<th>Rank</th>
<th>MA Population</th>
<th>Population Size Relative to New York</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>New York-Northern New Jersey-Long Island, NY-NJ-CT-PA</td>
<td>1.000</td>
</tr>
<tr>
<td>2</td>
<td>Los Angeles-Riverside-Orange County, CA</td>
<td>1.295</td>
</tr>
<tr>
<td>3</td>
<td>Chicago-Gary-Kenosha, IL-IN-WI</td>
<td>2.315</td>
</tr>
<tr>
<td>4</td>
<td>Washington-Baltimore, DC-MD-VA-WV</td>
<td>2.787</td>
</tr>
<tr>
<td>5</td>
<td>San Francisco-Oakland-San Jose, CA</td>
<td>3.012</td>
</tr>
<tr>
<td>6</td>
<td>Philadelphia-Wilmington-Atlantic City, PA-NJ-DE-MD</td>
<td>3.426</td>
</tr>
<tr>
<td>7</td>
<td>Boston-Worcester-Lawrence, MA-NH-ME-CT</td>
<td>3.643</td>
</tr>
<tr>
<td>8</td>
<td>Detroit-Ann Arbor-Flint, MI</td>
<td>3.885</td>
</tr>
<tr>
<td>9</td>
<td>Dallas-Fort Worth, TX</td>
<td>4.060</td>
</tr>
<tr>
<td>10</td>
<td>Houston-Galveston-Brazoria, TX</td>
<td>4.540</td>
</tr>
</tbody>
</table>

Note: S

Source: Census Bureau, 2000.
Citizens in large cities must devote part of their leisure time to nonproductive but work-related commuting.

The model, like Lucas and Rossi-Hansberg’s (2002) theory of the internal structure of cities, incorporates those two counteracting external forces. The current model does not explicitly model internal geographic heterogeneity of the city. Because in Lucas and Rossi-Hansberg (2002) citizens obtain the same utility over different locations, it is without loss of generality that citizens within a given city are considered identical. The main objective is to understand economic and population differences between cities, rather than within cities. The city is therefore not considered in isolation, but rather experiences population mobility from and to different cities. The main aim is to extend the work in this literature on the internal structure of cities and allow for competition between cities of different sizes. The space in which heterogeneous cities are considered is therefore the size space rather than a given geographical space.

Define an economy with local externalities $C$. Time is discrete and indexed by $t$. Let there be a set of locations (cities) $i = 1, \ldots, I$. Each city has a continuum population of size $S_i, t$, and the total, country-wide population size $S = \sum_{i} S_i, t$. All individuals are infinitely lived and can perform exactly one job. Let $A_i, t$ be the productivity parameter that reflects the technological advancement of city $i$ at time $t$. The law of motion of $A_i, t$ is $A_i, t = A_i, t / (1 + (1 / \mu_i, t))$. Each city experiences an exogenous technology shock $\mu_i, t$. Let $\mu_t$ denote the vector of shocks of all cities. The city-specific shock is symmetric and is identically and independently distributed with mean zero, and $\mu_i, t = 0$. On aggregate, there is no growth in productivity. This law of motion implies that $\ln(A_i, t)$ follows a unit root process. In empirical applications, the presence of a unit root often cannot be rejected. In the real business cycle literature, for example, using the Solow residual to measure TFP, the point estimates found on the persistence parameter in $A_i, t / (1 + (1 / \mu_i, t))$ cannot be rejected to be different from 1 (see, for example, Robert G. King and Sergio T. Rebelo, 1999).

Recent work by Rossi-Hansberg and Wright (2004) and Gilles Duranton (2002) has proposed different growth models that can explain Zipf’s law. Rossi-Hansberg and Wright (2004), for example, have shocks at the industry level. The implication is that while industry size is persistent over time, the size of a given city is not related to that of industries, as industries and workers can relocate each pe-

### Table 3—Pareto Efficient Regressions

<table>
<thead>
<tr>
<th></th>
<th>Estimates</th>
<th>s.e.</th>
<th>GI s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truncation point Estimate</td>
<td>$R^2$</td>
<td>N/S</td>
<td>City</td>
</tr>
<tr>
<td>135</td>
<td>Chattanooga (city), TN</td>
<td>21.099</td>
<td>1.354</td>
</tr>
<tr>
<td>19,383</td>
<td>Lyndhurst (CDP), NJ</td>
<td>20.648</td>
<td>1.314</td>
</tr>
<tr>
<td>6,592</td>
<td>Attalla (city), AL</td>
<td>18.588</td>
<td>1.125</td>
</tr>
<tr>
<td>1,378</td>
<td>Fullerton (city), NE</td>
<td>15.944</td>
<td>0.863</td>
</tr>
<tr>
<td>42</td>
<td>Paoli (town), CO</td>
<td>13.029</td>
<td>0.534</td>
</tr>
</tbody>
</table>

Notes: Dependent variable: Rank (ln). s.e. standard error; GI s.e. Gabaix-Ioannides (2003) corrected standard error ($\hat{a}_t (2/N)^{1/2}$).

Source: Census Bureau, 2000.
Evidence
All Cities

Citizens in large cities must devote part of their leisure time to nonproductive but work-related commuting.

The model, like Lucas and Rossi-Hansberg’s (2002) theory of the internal structure of cities, incorporates those two counteracting external forces. The current model does not explicitly model internal geographic heterogeneity of the city. Because in Lucas and Rossi-Hansberg (2002) citizens obtain the same utility over different locations, it is without loss of generality that citizens within a given city are considered identical. The main objective is to understand economic and population differences between cities, rather than within cities. The city is therefore not considered in isolation, but rather experiences population mobility from and to different cities. The main aim is to extend the work in this literature on the internal structure of cities and allow for competition between cities of different sizes. The space in which heterogeneous cities are considered is therefore the size space rather than a given geographical space.

Define an economy with local externalities $C$.

Time is discrete and indexed by $t$. Let there be a set of locations (cities) $\{1, \ldots, I\}$.

Each city has a continuum population of size $S_i$, and the total, country-wide population size $S$.

All individuals are infinitely lived and can perform exactly one job. Let $A_i, t$ be the productivity parameter that reflects the technological advancement of city $i$ at time $t$. The law of motion of $A_i, t$ is $A_i, t = A_{i, t-1} (1 + \eta_i, t)$.

Each city experiences an exogenous technology shock $\eta_i, t$. Let $\eta_t$ denote the vector of shocks of all cities. The city-specific shock is symmetric and is identically and independently distributed with mean zero, and $\eta_i, t = 0$.

On aggregate, there is no growth in productivity. This law of motion implies that $\ln(A_i, t)$ follows a unit root process. In empirical applications, the presence of a unit root often cannot be rejected. In the real business cycle literature, for example, using the Solow residual to measure TFP, the point estimates found on the persistence parameter $a$ in $A_i, t$ cannot be rejected to be different from 1 (see, for example, Robert G. King and Sergio T. Rebelo, 1999).

Recent work by Rossi-Hansberg and Wright (2004) and Gilles Duranton (2002) has proposed different growth models that can explain Zipf’s law. Rossi-Hansberg and Wright (2004), for example, have shocks at the industry level. The implication is that while industry size is persistent over time, the size of a given city is not related to that of industries, as industries and workers can relocate each other.

### Table 3—Pareto Coefficient Regressions

<table>
<thead>
<tr>
<th>Truncation point</th>
<th>City</th>
<th>Estimates</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>$S$</td>
<td>$\hat{K}$ (s.e.)</td>
<td>$\hat{a}$ (s.e.) (GI s.e.)</td>
<td>$R^2$</td>
</tr>
<tr>
<td>135</td>
<td>155,554 Chattanooga (city), TN</td>
<td>21.099</td>
<td>1.354</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.144)</td>
<td>(0.011) (0.165)</td>
<td></td>
</tr>
<tr>
<td>2,000</td>
<td>19,383 Lyndhurst (CDP), NJ</td>
<td>20.648</td>
<td>1.314</td>
<td>0.997</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.017)</td>
<td>(0.002) (0.042)</td>
<td></td>
</tr>
<tr>
<td>5,000</td>
<td>6,592 Attalla (city), AL</td>
<td>18.588</td>
<td>1.125</td>
<td>0.985</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.019)</td>
<td>(0.002) (0.023)</td>
<td></td>
</tr>
<tr>
<td>12,500</td>
<td>1,378 Fullerton (city), NE</td>
<td>15.944</td>
<td>0.863</td>
<td>0.961</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.014)</td>
<td>(0.002) (0.011)</td>
<td></td>
</tr>
<tr>
<td>25,000</td>
<td>42 Paoli (town), CO</td>
<td>13.029</td>
<td>0.534</td>
<td>0.860</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.010)</td>
<td>(0.001) (0.005)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Dependent variable: Rank (ln). s.e. standard error; GI s.e. Gabaix-Ioannides (2003) corrected standard error ($\hat{a}(2/N)^{1/2}$).

Source: Census Bureau, 2000.
From Population to Economics

- What drives population mobility?
  1. Geography: rivers, coasts, mountains, weather
  2. Amenities: Opera, externalities (+/-, (non-)pecuniary), ...
  3. Productivity Changes

- Citizen mobility in response to changes in prices: wages, housing prices, consumption prices,...

- Prices are determined in equilibrium

→ A general equilibrium theory of production across locations

∴ Objective: understand economic mechanisms (technology, preferences,...) from observing the population dynamics
From Population to Economics

- Local TFP $A_{i,t}$; law of motion: $A_{i,t} = A_{i,t-1}(1 + \sigma_{i,t})$ where $\sigma_{i,t}$ is zero mean i.i.d.
- Local externalities:
  - positive in production $a_+(S_{i,t}) (a'_+(S_{i,t}) > 0)$
  - negative (commuting) $a_-(S_{i,t}) (a'_-(S_{i,t}) < 0)$
- Identical firms in a competitive local labor market produce $y_{i,t} = A_{i,t} a_+(S_{i,t}) \Rightarrow$ wage is equal to marginal product
- Stock of land in each city is $H$; unit price of land is $p_{i,t}$ and individual consumption is $h_{i,t}$
- Preferences: $u(c, h, l) = c^\alpha h^\beta (1 - l)^{1-\alpha-\beta}$
- Perfect mobility across cities (no moving cost)

**Proposition**

*Under general conditions, city size satisfies Gibrat’s law: population growth is proportionate and the asymptotic size distribution is lognormal.*
What is a City?
What is a City?
What is a City?
What is a City?
What is a City?
Àrea Metropolitana de Barcelona
What is a City?

MSA, Place, County,...
What is a City?

Counties

The fact that Gibrat’s proposition is established concerning the population mobility of cities is a necessary requirement for an empirically consistent theory of the underlying economic activity. The second main purpose of this paper is to propose and solve an equilibrium model of local externalities where wages and prices guide citizens in their location decision. Consistent with proportionate growth and a log-normal size distribution, the model establishes a mechanism of local productivity shocks in the presence of local externalities and their effect, through worker mobility, on the population size distribution of cities.

Appendix A: The Size Distribution of Counties

We investigate the size distribution of counties. While counties may not necessarily be the right geographical unit that an economist is interested in, they do have the major advantage that the size distribution of counties comprises 100 percent of the U.S. population, i.e., 281 million in 2000. According to the Census, counties are described as the primary legal divisions of most states. For example, voting for most elections is organized at the county level. Most counties are functioning governmental units, whose powers and functions vary from state to state. Legal changes to county boundaries or names are typically infrequent. In 2000, there were 3,141 counties in the United States covering the entire population. The ten largest are listed in Table A-1. The largest, Los Angeles County, California, had 9.5 million inhabitants and the smallest, Loving County, Texas, 67 inhabitants. The sample mean (in ln, standard error in brackets) is 10.22 (0.02) and the standard deviation is 1.41.

In Figure A-1 we plot the size distribution, together with the normal density. The size empirical density is remarkably similar to the normal. There is somewhat more mass near the mode, and the distribution may be slightly skewed, but the distribution of county size is nonetheless surprisingly close to lognormal.

Table A-1—The Largest Counties in the United States

<table>
<thead>
<tr>
<th>Rank</th>
<th>City</th>
<th>Population $S$</th>
<th>$S_{LA}/S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Los Angeles County, CA</td>
<td>9,519,338</td>
<td>1.000</td>
</tr>
<tr>
<td>2</td>
<td>Cook County, IL</td>
<td>5,376,741</td>
<td>1.770</td>
</tr>
<tr>
<td>3</td>
<td>Harris County, TX</td>
<td>3,400,578</td>
<td>2.799</td>
</tr>
<tr>
<td>4</td>
<td>Maricopa County, AZ</td>
<td>3,072,149</td>
<td>3.099</td>
</tr>
<tr>
<td>5</td>
<td>Orange County, CA</td>
<td>2,846,289</td>
<td>3.344</td>
</tr>
<tr>
<td>6</td>
<td>San Diego County, CA</td>
<td>2,813,833</td>
<td>3.383</td>
</tr>
<tr>
<td>7</td>
<td>Kings County, NY</td>
<td>2,465,326</td>
<td>3.861</td>
</tr>
<tr>
<td>8</td>
<td>Miami-Dade County, FL</td>
<td>2,253,362</td>
<td>4.225</td>
</tr>
<tr>
<td>9</td>
<td>Queens County, NY</td>
<td>2,229,379</td>
<td>4.269</td>
</tr>
<tr>
<td>10</td>
<td>Dallas County, TX</td>
<td>2,218,899</td>
<td>4.290</td>
</tr>
</tbody>
</table>

Note: $S_{LA}/S$ denotes the ratio of population size relative to Los Angeles.

Source: Census Bureau, 2000.
What is a City?

Counties

The fact that Gibrat’s proposition is established concerning the population mobility of cities is a necessary requirement for an empirically consistent theory of the underlying economic activity. The second main purpose of this paper is to propose and solve an equilibrium model of local externalities where wages and prices guide citizens in their location decision. Consistent with proportionate growth and a log-normal size distribution, the model establishes a mechanism of local productivity shocks in the presence of local externalities and their effect, through worker mobility, on the population size distribution of cities.

Appendix A: The Size Distribution of Counties

We investigate the size distribution of counties. While counties may not necessarily be the right geographical unit that an economist is interested in, they do have the major advantage that the size distribution of counties comprises 100 percent of the U.S. population, i.e., 281 million in 2000. According to the Census, counties are described as the primary legal divisions of most states. For example, voting for most elections is organized at the county level. Most counties are functioning governmental units, whose powers and functions vary from state to state. Legal changes to county boundaries or names are typically infrequent.

In 2000, there were 3,141 counties in the United States covering the entire population. The ten largest are listed in Table A-1. The largest, Los Angeles County, California, had 9.5 million inhabitants and the smallest, Loving County, Texas, 67 inhabitants. The sample mean (in ln, standard error in brackets) is \( \ln 10.22 \pm 0.02 \) and the standard deviation is \( 1.41 \).

In Figure A-1 we plot the size distribution, together with the normal density function. The size empirical density is remarkably similar to the normal. There is somewhat more mass near the mode, and the distribution may be slightly skewed, but the distribution of county size is nonetheless surprisingly close to lognormal.

Table A-1—The Largest Counties in the United States

<table>
<thead>
<tr>
<th>Rank</th>
<th>City, State</th>
<th>Population</th>
<th>LA/S</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Los Angeles, CA</td>
<td>9,519,338</td>
<td>1.000</td>
</tr>
<tr>
<td>2</td>
<td>Cook, IL</td>
<td>5,376,741</td>
<td>1.770</td>
</tr>
<tr>
<td>3</td>
<td>Harris, TX</td>
<td>3,400,578</td>
<td>2.799</td>
</tr>
<tr>
<td>4</td>
<td>Maricopa, AZ</td>
<td>3,072,149</td>
<td>3.099</td>
</tr>
<tr>
<td>5</td>
<td>Orange, CA</td>
<td>2,846,289</td>
<td>3.344</td>
</tr>
<tr>
<td>6</td>
<td>San Diego, CA</td>
<td>2,813,833</td>
<td>3.383</td>
</tr>
<tr>
<td>7</td>
<td>Kings, NY</td>
<td>2,465,326</td>
<td>3.861</td>
</tr>
<tr>
<td>8</td>
<td>Miami-Dade, FL</td>
<td>2,253,362</td>
<td>4.225</td>
</tr>
<tr>
<td>9</td>
<td>Queens, NY</td>
<td>2,229,379</td>
<td>4.269</td>
</tr>
<tr>
<td>10</td>
<td>Dallas, TX</td>
<td>2,218,899</td>
<td>4.290</td>
</tr>
</tbody>
</table>

Note: \( \text{LA/S denotes the ratio of population size relative to Los Angeles.} \)

Source: Census Bureau, 2000.

Appendix B: Additional Statistics of City Growth

Table B-1—Descriptive Statistics of City Growth

<table>
<thead>
<tr>
<th>Range of cities</th>
<th>Growth rate (non-normalized)</th>
<th>N</th>
<th>mean</th>
<th>stdev</th>
<th>IQR (Q3/Q1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>19,361</td>
<td>0.103</td>
<td>0.729</td>
<td>0.199</td>
<td></td>
</tr>
<tr>
<td>Top 100</td>
<td>100</td>
<td>0.108</td>
<td>0.158</td>
<td>0.154</td>
<td></td>
</tr>
<tr>
<td>Bottom 100</td>
<td>100</td>
<td>0.127</td>
<td>0.671</td>
<td>0.493</td>
<td></td>
</tr>
<tr>
<td>P10 to P90</td>
<td>15,488</td>
<td>0.106</td>
<td>0.786</td>
<td>0.191</td>
<td></td>
</tr>
</tbody>
</table>

What is a City?

Constructing cities

Holmes and Lee: a unit consists of a $6 \times 6$ miles area

Fig. 3.1 Map of grid lines for six-by-six squares in the vicinity of New York City
Outline

I  Zipf’s and Gibrat’s law
II  Spatial Sorting
III  Taxation
Spatial Sorting
The Urban Wage Premium
Spatial Sorting
The Urban Wage Premium

The graph shows a scatter plot with the average log wage on the y-axis and the log population on the x-axis. There is a positive correlation between the average log wage and the log population, indicated by a linear trend line that slopes upwards from left to right.
**Spatial Sorting**

**The Urban Wage Premium**

- The elasticity of average wage with respect to city size is 4.2%
- Big differences:

<table>
<thead>
<tr>
<th></th>
<th>Population</th>
<th>Wage</th>
<th>Wage Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York</td>
<td>19 million</td>
<td>897</td>
<td>1.22</td>
</tr>
<tr>
<td>Janesville, WI</td>
<td>160,000</td>
<td>735</td>
<td>1.00</td>
</tr>
</tbody>
</table>

⇒ Explanations?

1. Amenities
2. Cost of Living
3. Sorting
The elasticity of average wage with respect to city size is 4.2%.

Big differences:

<table>
<thead>
<tr>
<th>City</th>
<th>Population</th>
<th>Wage</th>
<th>Wage Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York</td>
<td>19 million</td>
<td>897</td>
<td>1.22</td>
</tr>
<tr>
<td>Janesville, WI</td>
<td>160,000</td>
<td>735</td>
<td>1.00</td>
</tr>
</tbody>
</table>

⇒ Explanations?

1. Amenities
2. Cost of Living
3. Sorting
Spatial Sorting
The Urban Wage Premium

- The elasticity of average wage with respect to city size is 4.2%
- Big differences:

<table>
<thead>
<tr>
<th></th>
<th>Population</th>
<th>Wage</th>
<th>Wage Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York</td>
<td>19 million</td>
<td>897</td>
<td>1.22</td>
</tr>
<tr>
<td>Janesville, WI</td>
<td>160,000</td>
<td>735</td>
<td>1.00</td>
</tr>
</tbody>
</table>

⇒ Explanations?
1. Amenities
2. Cost of Living
3. Sorting
SORTING IN TEAMS
PRODUCTION AND COMPLEMENTARITIES
SORTING IN TEAMS
PRODUCTION AND COMPLEMENTARITIES
SORTING IN TEAMS
PRODUCTION AND COMPLEMENTARITIES
SORTING IN TEAMS

PRODUCTION AND COMPLEMENTARITIES
SORTING IN TEAMS
PRODUCTION AND COMPLEMENTARITIES
SORTING IN TEAMS
PRODUCTION AND COMPLEMENTARITIES
Sorting in Teams
Production and Complementarities
SORTING IN TEAMS
PRODUCTION AND COMPLEMENTARITIES
Sorting in Teams
Production and Complementarities
Sorting in Teams
Production and Complementarities
City as a Team
Production and Complementarities

- From team to city
- Aggregate production technology with specific complementarities
- Additional economic forces: housing prices
City as a team
Production and Complementarities

- From team to city
- Aggregate production technology w/ specific complementarities
- Additional economic forces: housing prices
- Objective: derive skill complementarities from choice of citizens where to live/work

→ Spatial Sorting
City as a Team
Spatial Sorting
The model

- \( J \) locations (cities) \( j \in \mathcal{J} = \{1, \ldots, J\} \)
- Fixed amount of land (housing) \( H_j \)
• Citizens (workers) with heterogenous skills $x_i$
• Preferences over consumption and housing (price $p$):

$$u(c, h) = c^{1-\alpha} h^\alpha$$

• Worker mobility $\Rightarrow$ utility equalization across cities:

$$u(c_{ij}, h_{ij}) = u(c_{ij'}, h_{ij'}), \quad \forall j' \neq j$$
Technology

- Cities differ exogenously in TFP \( A_j \)
- Representative firm in city \( j \) produces

\[
A_j F(m_{1j}, \ldots, m_{ij})
\]

\( m_{ij} \): employment level of skill \( i \); given wages \( w_{ij} \)
Technology: Nested CES
3 skill types ⇒ 5 configurations

0. Benchmark CES:

\[ A_j F = A_j \left( m_{1j}^\gamma y_1 + m_{2j}^\gamma y_2 + m_{3j}^\gamma y_3 \right)^\beta \quad \gamma \in [0, 1], \beta > 0 \]
Technology: Nested CES

3 skill types ⇒ 5 configurations

0. Benchmark CES:

\[ A_j F = A_j \left( m_1^\gamma y_1 + m_2^\gamma y_2 + m_3^\gamma y_3 \right)^\beta \quad \gamma \in [0, 1], \beta > 0 \]

1. Extreme-Skill Complementarity

\[ A_j F = A_j \left[ m_2^\gamma y_2 + (m_1^\gamma y_1 + m_3^\gamma y_3)^\lambda \right]^\beta \]

A. \( \lambda > 1 \): skills 1 and 3 are (relative) complements;
B. \( \lambda < 1 \): skills 1 and 3 are (relative) substitutes;
C. \( \lambda = 1 \): CES
Technology: Nested CES

3 skill types ⇒ 5 configurations

0. Benchmark CES:

\[ A_j F = A_j \left( m_1^\gamma y_1 + m_2^\gamma y_2 + m_3^\gamma y_3 \right)^\beta \quad \gamma \in [0, 1], \beta > 0 \]

1. Extreme-Skill Complementarity

\[ A_j F = A_j \left[ m_2^\gamma y_2 + (m_1^\gamma y_1 + m_3^\gamma y_3)^\lambda \right]^\beta \]

- A. \( \lambda > 1 \): skills 1 and 3 are (relative) complements;
- B. \( \lambda < 1 \): skills 1 and 3 are (relative) substitutes;
- C. \( \lambda = 1 \): CES

2. Top-Skill Complementarity

\[ A_j F = A_j \left[ m_1^\gamma y_1 + (m_2^\gamma y_2 + m_3^\gamma y_3)^\lambda \right]^\beta \]
Market clearing

- Housing market: \( \sum_{i=1}^{l} h_{ij} m_{ij} = H_j \)
- Labour market: \( \sum_{j=1}^{J} m_{ij} = M_i \) \( (M_i: \text{total \# of skill } i) \)
- City population: \( S_j = \sum_{i=1}^{l} m_{ij} \)
- Two types of cities, \( C_1, C_2 \) of each type
Citizen’s problem

- Optimal consumption
  \[ c_{ij}^* = (1 - \alpha)w_{ij} \quad \text{and} \quad h_{ij}^* = \alpha \frac{w_{ij}}{p_j} \]

- Indirect utility function
  \[ U_i = \alpha^\alpha (1 - \alpha)^{1-\alpha} \frac{w_{ij}}{p_j^\alpha} \]

⇒ From mobility, utility equalization:
  \[ \frac{w_{i1}}{p_1^\alpha} = \frac{w_{i2}}{p_2^\alpha} \]
Main Results

Theorem 1. City Size and TFP

The more productive city is larger, $S_1 > S_2$
Main Results

Theorem 1. City Size and TFP
The more productive city is larger, $S_1 > S_2$

Theorem 2. Extreme-Skill Complementarity and Thick Tails
The skill distribution in the larger city has thicker tails
Main Results

Theorem 1. City Size and TFP
The more productive city is larger, $S_1 > S_2$

Theorem 2. Extreme-Skill Complementarity and Thick Tails
The skill distribution in the larger city has thicker tails

Theorem 3. Top-Skill Complementarity and FOSD
The skill distribution in the larger city first-order stoch. dominates
Main Results

Mechanism: skill complementarity also in small cities, but demand for extreme skills is higher in big cities due to TFP ($A_j$)
Main Results

Mechanism: skill complementarity also in small cities, but demand for extreme skills is higher in big cities due to TFP ($A_j$)

Corollary 1. CES technology
If $\lambda = 1$, then the skill distribution across cities is identical

Corollary 2. Extreme-Skill Substitutability and Thin Tails
The skill distribution in the larger city has thinner tails
Main Results
5 technologies → 5 distributions

1. Extreme-Skill Complementarity ⇒ thick tails
2. Extreme-Skill Substitutability ⇒ thin tails
3. Top-Skill Complementarity ⇒ FOSD of big cities
4. Top-Skill Substitutability ⇒ FOSD of small cities
5. Constant Elasticity (CES) ⇒ identical distributions
Empirical evidence

Use theory to obtain a measure for skills

\[ u_i = \alpha \frac{1}{1 - \alpha} \frac{w_{ij}}{p_j} \alpha \]

Need to observe:
- wage distribution \( w_{ij} \) by city
- housing price level \( p_j \)
- budget share of housing \( \hat{\alpha} = 0.24 \) from Davis and Ortalo-Magne (RED 2010)
**Empirical evidence**

- Use theory to obtain a measure for skills

\[ U_i = \alpha^\alpha (1 - \alpha)^{1-\alpha} \frac{w_{ij}}{p_j^\alpha} \]

- Need to observe:
  - wage distribution \( w_{ij} \) by city
  - housing price level \( p_j \)
  - budget share of housing \( \alpha \)
    \( \hat{\alpha} = 0.24 \) from Davis and Ortalo-Magné (RED 2010)
log wage population < 1m > 2.5m

10th percentile: pop < 1m = 5.93, pop > 2.5m = 5.99, diff = 0.065*** (0.007)
90th percentile: pop < 1m = 7.36, pop > 2.5m = 7.56, diff = 0.198*** (0.007)
Housing prices

- American Community Survey (ACS) 2009
- Rental prices (robust: sales)
- Hedonic price schedule: to obtain housing price index

⇒ Skill measure: \( \frac{w_i}{p_i^\alpha} \)
Skills and city size

Skill measure: $\frac{w_i}{p_i^\alpha}$

10th percentile: pop < 1m = 5.44, pop > 2.5m = 5.36, diff = -0.074*** (0.006)
90th percentile: pop < 1m = 6.86, pop > 2.5m = 6.99, diff = 0.132*** (0.009)
Skills and city size

1. Constant mean: housing cost increases $4 \times$ faster than wages
   \[ \Rightarrow 1.169^{0.24} = 1.038 \approx 1.042 \]

2. Variance increases in city size

\[ \therefore \text{Urban Wage Premium: not spatial sorting, but housing prices} \]
Skills and city size

1. Constant mean: housing cost increases $4 \times$ faster than wages
   $\Rightarrow 1.169^{0.24} = 1.038 \approx 1.042$

2. Variance increases in city size

   $\therefore$ Urban Wage Premium: not spatial sorting, but housing prices
   $\therefore$ Skill distribution thick tails $\rightarrow$ extreme-skill complementarity

   $$A_j F = A_j \left[ m_2^\gamma y_2 + (m_1^\gamma y_1 + m_3^\gamma y_3)^\lambda \right]^{\beta}, \quad \lambda > 1$$
Skills and city size

1. Constant mean: housing cost increases $4 \times$ faster than wages
   \[ \Rightarrow 1.169^{0.24} = 1.038 \approx 1.042 \]

2. Variance increases in city size

   \[ \therefore \text{Urban Wage Premium: not spatial sorting, but housing prices} \]
   \[ \therefore \text{Skill distribution thick tails} \rightarrow \text{extreme-skill complementarity} \]

   \[
   A_j F = A_j \left[ m_2^\gamma y_2 + (m_1^\gamma y_1 + m_3^\gamma y_3)^\lambda \right]^\beta, \quad \lambda > 1
   \]

   \[ \rightarrow \text{high skilled workers need low-skilled services for production} \]
   \begin{itemize}
   
   \item administrative/sales help
   \item household help and child care
   \item food services, restaurants,...
   
   \end{itemize}
Robustness: Observables

- Our measure of skills: price based (wages and housing price)
- Includes everything: observables and unobservables
- 2/3 of wages: unobservables (non-cognitive skills,...)

→ Thick tails also for observables?
**Education: A Direct Measure of Skill**

- No high school
- High school diploma
- Bachelor's and more

<table>
<thead>
<tr>
<th>Type</th>
<th>Population &lt; 1m</th>
<th>Population &gt; 2.5m</th>
</tr>
</thead>
<tbody>
<tr>
<td>10th percentile:</td>
<td>-0.61</td>
<td>-0.65</td>
</tr>
<tr>
<td>90th percentile:</td>
<td>0.64</td>
<td>0.67</td>
</tr>
<tr>
<td>Diff:</td>
<td>-0.046***</td>
<td>0.032***</td>
</tr>
<tr>
<td>(0.007)</td>
<td>(0.008)</td>
<td></td>
</tr>
</tbody>
</table>

Kurtosis (H0: = 3):
- Population < 1m = 2.99
- Population > 2.5m = 2.92***

Wage data: cps 2009, obs pop < 1m = 25584, obs pop > 2.5m = 34999

Dep. var.: \( l_{resid2utility20grade92} \) = Residual Skill(utility20), not predicted by grade92

6 Jan 2013, 23:12:42
Occupation

- Density of occupation for low, middle, and high density areas, with bars indicating population size:
  - Low population < 1m: low density
  - Middle population > 2.5m: middle density
  - High population: high density

- PDF of residual skill, controlled for occupation:
  - 10th percentile: pop < 1m = -0.55, pop > 2.5m = -0.59, diff = -0.042*** (0.006)
  - 90th percentile: pop < 1m = 0.56, pop > 2.5m = 0.60, diff = 0.040*** (0.007)
Industrial Composition

10th percentile: pop < 1m = -0.63, pop > 2.5m = -0.69, diff = -0.053*** (0.006)
90th percentile: pop < 1m = 0.66, pop > 2.5m = 0.74, diff = 0.074*** (0.008)
**Migration**

---

**Foreign Born**

- 10th percentile: pop < 1m = 5.23, pop > 2.5m = 5.14, diff = -0.085*** (0.017)
- 90th percentile: pop < 1m = 6.61, pop > 2.5m = 6.70, diff = 0.083** (0.046)

**Dep. var.: lutility20 = Skill: cbsa rentindex (ACS 2009)**

---

**Natives**

- 10th percentile: pop < 1m = 5.47, pop > 2.5m = 5.45, diff = -0.014** (0.007)
- 90th percentile: pop < 1m = 6.87, pop > 2.5m = 7.02, diff = 0.151*** (0.010)

**Dep. var.: lutility20 = Skill: cbsa rentindex (ACS 2009)**

---

**Kurtosis (H0: =3)**

- pop < 1m = 3.11, pop > 2.5m = 2.85**

---

**Wage data: cps 2009, obs pop < 1m = 1371, obs pop > 2.5m = 4402**
**AGE**

- **20-29 year old**
  - 10th percentile: pop < 1m = 5.32, pop > 2.5m = 5.27, diff = -0.051*** (0.012)
  - 90th percentile: pop < 1m = 6.48, pop > 2.5m = 6.57, diff = 0.090*** (0.018)
  - Kurtosis (H0: =3): pop < 1m = 3.29***, pop > 2.5m = 2.87**
  - Wage data: cps 2009, obs pop < 1m = 4806, obs pop > 2.5m = 6591
- **30-39 year old**
  - 10th percentile: pop < 1m = 5.48, pop > 2.5m = 5.38, diff = -0.092*** (0.014)
  - 90th percentile: pop < 1m = 6.84, pop > 2.5m = 6.97, diff = 0.131*** (0.019)
  - Kurtosis (H0: =3): pop < 1m = 2.72***, pop > 2.5m = 2.43***
- **40-49 year old**
  - 10th percentile: pop < 1m = 5.51, pop > 2.5m = 5.45, diff = -0.057*** (0.014)
  - 90th percentile: pop < 1m = 6.95, pop > 2.5m = 7.11, diff = 0.158*** (0.018)
  - Kurtosis (H0: =3): pop < 1m = 2.59***, pop > 2.5m = 2.29***
- **50-59 year old**
  - 10th percentile: pop < 1m = 5.53, pop > 2.5m = 5.45, diff = -0.073*** (0.015)
  - 90th percentile: pop < 1m = 6.99, pop > 2.5m = 7.09, diff = 0.101*** (0.019)
  - Kurtosis (H0: =3): pop < 1m = 2.55***, pop > 2.5m = 2.36***

Wage data: cps 2009, obs pop < 1m = 4806, obs pop > 2.5m = 6591
Dep. var.: lutility20 = Skill: cbsa rentindex (ACS 2009)
3 Jan 2013, 21:00:13
**Decomposing the skill distributions**

**Small vs. big cities**

<table>
<thead>
<tr>
<th></th>
<th>10% Quantile</th>
<th>90% Quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observed Quantiles:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Large cities</td>
<td>5.365 (0.004)***</td>
<td>6.994 (0.006)***</td>
</tr>
<tr>
<td>- Small cities</td>
<td>5.439 (0.005)***</td>
<td>6.862 (0.007)***</td>
</tr>
<tr>
<td>- Difference</td>
<td>-0.074 (0.006)***</td>
<td>0.132 (0.009)***</td>
</tr>
</tbody>
</table>

*Firpo, Fortin, Lemieux (2009)*

**Predicted Quantiles:**

<table>
<thead>
<tr>
<th></th>
<th>10% Quantile</th>
<th>90% Quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Large cities</td>
<td>5.387 (0.005)***</td>
<td>7.022 (0.005)***</td>
</tr>
<tr>
<td>- Small cities</td>
<td>5.454 (0.004)***</td>
<td>6.878 (0.008)***</td>
</tr>
<tr>
<td>- Difference</td>
<td>-0.068 (0.007)***</td>
<td>0.144 (0.009)***</td>
</tr>
</tbody>
</table>

Explained by observables:

- Education (16 categories) 0.003 (0.002) ** 0.052 (0.002) ***
- Occupation (22 categories) 0.004 (0.002) * 0.025 (0.003) ***
- Industry (51 categories) -0.001 (0.002) 0.013 (0.002) ***
- Race (4 groups) -0.004 (0.001) *** -0.015 (0.001) ***
- Sex -0.001 (0.001) * -0.002 (0.001) *
- Foreign born -0.020 (0.002) *** -0.004 (0.001) ***
- Age (2nd order polynomial) 0.000 (0.001) -0.002 (0.001) *

Total explained by observables -0.018 (0.004) *** 0.067 (0.005) ***

Not explained by observables -0.049 (0.006) *** 0.077 (0.008) ***

*Chernozhukov, Fernández-Val, Melly (2012)*

**Predicted Quantile difference**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.068 (0.006)</td>
<td>0.113 (0.009)</td>
<td></td>
</tr>
</tbody>
</table>

Explained by observables -0.019 (0.004) 0.064 (0.005)

Not explained by observables -0.050 (0.007) 0.049 (0.007)
SORTING WITHIN CITIES
DETROIT

Legend
Rental Index 2009
- 0.5 - 0.7
- 0.7 - 0.8
- 0.8 - 0.9
- 0.9 - 1.0
- 1.0 - 1.1
- 1.1 - 1.2
- 1.2 - 1.3
- 1.3 - 1.4
I Zipf’s and Gibrat’s law
II Spatial Sorting
III Taxation
Federal Taxes affect same skill workers differentially in cities:

- Urban Wage Premium
- Progressive Taxation
Federal Taxes affect same skill workers differentially in cities:
  - Urban Wage Premium
  - Progressive Taxation

Average tax rate: 3% points difference at median:

<table>
<thead>
<tr>
<th>Population</th>
<th>Wage level</th>
<th>Avg. Tax Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York</td>
<td>19 million</td>
<td>1.22</td>
</tr>
<tr>
<td>Janesville, WI</td>
<td>160,000</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Income Taxation in Local Labor Markets

- Federal Taxes affect **same skill** workers differentially in cities:
  - Urban Wage Premium
  - Progressive Taxation

- Average tax rate: 3% points difference *at median*:

<table>
<thead>
<tr>
<th>Population</th>
<th>Wage level</th>
<th>Avg. Tax Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York</td>
<td>19 million</td>
<td>1.22</td>
</tr>
<tr>
<td>Janesville, WI</td>
<td>160,000</td>
<td>1.00</td>
</tr>
</tbody>
</table>

- Due to mobility: no redistribution! Same skills, same utility
- Policy Question: what is **optimal spatial taxation policy**?
Model

- $J$ cities, with TFP $A_j$; Identical agents; Output: $A_j l_j^\gamma$
- Amenities: $\varepsilon_j \rightarrow u(c, h) = (1 + \varepsilon_j)c^{1-\alpha}h^\alpha$
- Mobility: $u(c_j, h_j) = u(c_{j'}, h_{j'})$, $\forall j, j'$
- Tax schedule
  \[ \tilde{w}_j = \lambda w_j^{1-\tau} \]

- average tax rate: $\lambda w_j^{1-\tau}$
- marginal tax rate $\lambda(1 - \tau)w_j^{1-\tau}$
- $\tau = 0$: proportional; $\tau > 0$: progressive; $\tau < 0$: regressive
- US, estimated $\tau \approx 0.12$
Empirical Results

Parametrization

- Production: $\gamma = 1$ output $A_j l_j$

- Tax schedule: $\tau = 0.12, \lambda = 0.752$ (OECD calculator)

- Housing Exp. 24% (Davis, Ortalo-Magné, 2009)
  $\Rightarrow \alpha = \frac{0.24}{\lambda} = 0.319$
• TFP from average wages and labor force:

\[ A_j = \frac{w_j l_j^{1-\gamma}}{\gamma}, \quad \forall j. \]

• Amenities from mobility (utility equalization):

\[ 1 + \varepsilon_j = \frac{l_j^\alpha w_1^{(1-\alpha)(1-\tau US)}}{l_1^\alpha w_j^{(1-\alpha)(1-\tau US)}}, \quad \forall j. \]

• Revenue neutrality \( \rightarrow \) fixes \( \lambda \)
Optimal Tax Schedule?

- TFP from average wages and labor force:
  \[ A_j = \frac{w_j l_j^{1-\gamma}}{\gamma}, \quad \forall j. \]

- Amenities from mobility (utility equalization):
  \[ 1 + \varepsilon_j = \frac{l_j^\alpha w_1^{(1-\alpha)(1-\tau^{US})}}{l_1^\alpha w_j^{(1-\alpha)(1-\tau^{US})}} \]

- Revenue neutrality \( \rightarrow \) fixes \( \lambda \)

\[ \Rightarrow \forall \tau, \text{ new } l_j, u_j: \text{ search grid for } \tau \text{ that maximizes } u \]
**Optimal Tax Schedule**

\[ \tau^* = 9\% \]
Tax Schedules
Actual vs. Optimal

![Graph showing tax rates and relative wages for benchmark and optimal scenarios.]

- Tax rates: 0.15, 0.2, 0.25, 0.3, 0.4, 0.6, 0.8, 1, 1.2, 1.4
- Relative wages: 0.4, 0.6, 0.8, 1, 1.2, 1.4

- Benchmark line
- Optimal line
Simulation

Change in Labor Force – Productivity

![Graph showing the relationship between change in labor force and productivity across various cities in the U.S.](image)
Simulation

Change in Labor Force – Amenities

- Amarillo, TX
- Athens-Clark County, GA
- Bowling Green, KY
- Bridgeport-Stamford-Norwalk, CT
- Brownsville-Harlingen, TX
- Chicago-Naperville-Joliet, IL-IN-WI
- Los Angeles-Long Beach-Santa Ana, CA
- Ocean City, NJ
- Saginaw-Saginaw Township North, MI
- San Francisco-Oakland-Fremont, CA
- San Jose-Sunnyvale-Santa Clara, CA

Change in Labor Force (%)

-4 -2 0 2

Epsilon

Places represented on the graph include Amarillo, TX; Athens-Clark County, GA; Bowling Green, KY; Bridgeport-Stamford-Norwalk, CT; Brownsville-Harlingen, TX; Chicago-Naperville-Joliet, IL-IN-WI; Los Angeles-Long Beach-Santa Ana, CA; Ocean City, NJ; Saginaw-Saginaw Township North, MI; San Francisco-Oakland-Fremont, CA; and San Jose-Sunnyvale-Santa Clara, CA.
SIMULATION
CHANGE IN AFTER-TAX WAGES

Productivity

Change in After Tax Wages (%)

-2
-1
0
1
1.2
1.4

Amarillo, TX
Athens-Clark County, GA
Bowling Green, KY
Bridgeport-Stamford-Norwalk, CT
Brownsville-Harlingen, TX
Chicago-Naperville-Joliet, IL-IN-WI
Los Angeles-Long Beach-Santa Ana, CA
New York-Northern New Jersey-Long Island, NY-NJ-PA
Ocean City, NJ
Saginaw-Saginaw Township North, MI
San Francisco-Oakland-Fremont, CA
San Jose-Sunnyvale-Santa Clara, CA
SIMULATION

CHANGE IN HOUSING PRICES

- Amarillo, TX
- Athens-Clark County, GA
- Bowling Green, KY
- Bridgeport-Stamford-Norwalk, CT
- Brownsville-Harlingen, TX
- Chicago-Naperville-Joliet, IL-IN-WI
- Los Angeles-Long Beach-Santa Ana, CA
- New York-Northern New Jersey-Long Island, NY-NJ-PA
- Ocean City, NJ
- Saginaw-Saginaw Township North, MI
- San Francisco-Oakland-Fremont, CA
- San Jose-Sunnyvale-Santa Clara, CA

ChangeP

Productivity
# Outcomes for Selected Cities

<table>
<thead>
<tr>
<th>MSA</th>
<th>A</th>
<th>ε</th>
<th>Δl</th>
<th>%Δp</th>
<th>%Δc</th>
<th>%Δh</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Highest A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bridgeport-Stamford-Norwalk, CT</td>
<td>1.38</td>
<td>-0.16</td>
<td>1.62</td>
<td>2.39</td>
<td>0.76</td>
<td>-1.60</td>
</tr>
<tr>
<td>San Jose-Sunnyvale-Santa Clara, CA</td>
<td>1.36</td>
<td>0.14</td>
<td>1.55</td>
<td>2.28</td>
<td>0.72</td>
<td>-1.52</td>
</tr>
<tr>
<td>San Francisco-Oakland-Fremont, CA</td>
<td>1.35</td>
<td>0.44</td>
<td>1.52</td>
<td>2.24</td>
<td>0.71</td>
<td>-1.50</td>
</tr>
<tr>
<td><strong>Lowest A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brownsville-Harlingen, TX</td>
<td>0.53</td>
<td>0.00</td>
<td>-2.97</td>
<td>-4.32</td>
<td>-1.40</td>
<td>3.06</td>
</tr>
<tr>
<td>Amarillo, TX</td>
<td>0.49</td>
<td>-0.02</td>
<td>-3.31</td>
<td>-4.82</td>
<td>-1.56</td>
<td>3.42</td>
</tr>
<tr>
<td>Bowling Green, KY</td>
<td>0.46</td>
<td>-0.26</td>
<td>-3.65</td>
<td>-5.31</td>
<td>-1.72</td>
<td>3.79</td>
</tr>
<tr>
<td><strong>Highest ε</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New York-Northern New Jersey-Long Island</td>
<td>1.17</td>
<td>1.45</td>
<td>0.83</td>
<td>1.22</td>
<td>0.39</td>
<td>-0.82</td>
</tr>
<tr>
<td>Los Angeles-Long Beach-Santa Ana, CA</td>
<td>1.02</td>
<td>1.37</td>
<td>0.16</td>
<td>0.24</td>
<td>0.08</td>
<td>-0.16</td>
</tr>
<tr>
<td>Chicago-Naperville-Joliet, IL-IN-WI</td>
<td>1.06</td>
<td>1.07</td>
<td>0.35</td>
<td>0.52</td>
<td>0.17</td>
<td>-0.35</td>
</tr>
<tr>
<td><strong>Lowest ε</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saginaw-Saginaw Township North, MI</td>
<td>1.17</td>
<td>-0.46</td>
<td>0.81</td>
<td>1.19</td>
<td>0.38</td>
<td>-0.80</td>
</tr>
<tr>
<td>Athens-Clark County, GA</td>
<td>1.04</td>
<td>-0.53</td>
<td>0.27</td>
<td>0.40</td>
<td>0.13</td>
<td>-0.27</td>
</tr>
<tr>
<td>Ocean City, NJ</td>
<td>1.12</td>
<td>-0.63</td>
<td>0.62</td>
<td>0.92</td>
<td>0.29</td>
<td>-0.62</td>
</tr>
</tbody>
</table>
Simulation

\( c/h \) substitution

Amarillo, TX
Athens-Clark County, GA
Bowling Green, KY
Bridgeport-Stamford-Norwalk, CT
Brownsville-Harlingen, TX
Chicago-Naperville-Joliet, IL-IN-WI
Los Angeles-Long Beach-Santa Ana, CA
New York-Northern New Jersey-Long Island, NY-NJ-PA
Ocean City, NJ
San Francisco-Oakland-Fremont, CA
San Jose-Sunnyvale-Santa Clara, CA

Change in Goods Consumption (%) vs. Change in Housing Consumption

Brownsville-Harlingen, TX
Amarillo, TX
Bowling Green, KY

-2
-1
0
1
2
3
4

-2
-1
0
1
2
3
4

Change in Goods Consumption (%)
Change in Housing Consumption
**Aggregate Outcomes**

\[ \alpha = 0.319, \gamma = 1, \tau^* = 0.067 \]

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>%Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output gain</td>
<td>1.02</td>
</tr>
<tr>
<td>Population in 5 largest cities</td>
<td>0.59</td>
</tr>
<tr>
<td>Average housing prices</td>
<td>1.25</td>
</tr>
</tbody>
</table>
### Sensitivity

\[
\alpha = 0.24, \gamma = 1 \quad \alpha = 0.3191, \gamma = 1.2
\]

\[
\tau^* = -0.0082 \quad \tau^* = -0.0834
\]

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>%Δ</th>
<th>%Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output gain</td>
<td>8.86</td>
<td>20.30</td>
</tr>
<tr>
<td>Population in 5 largest cities</td>
<td>4.91</td>
<td>9.63</td>
</tr>
<tr>
<td>Average housing prices</td>
<td>10.36</td>
<td>23.39</td>
</tr>
</tbody>
</table>
1. Zipf’s law and Gibrat’s law
   - Puzzle resolved
Concluding Remarks
Economics and the City

1. Zipf’s law and Gibrat’s law
   - Puzzle resolved

2. There is Spatial Sorting
   - Thick tails $\rightarrow$ bigger inequality in big cities
   - Extreme-skill compl.: Urban wage premium not due to skills
     $\rightarrow$ increasing over time $+$ urbanization $\uparrow \Rightarrow$ inequality $\uparrow$
**Concluding Remarks**

**Economics and the City**

1. Zipf’s law and Gibrat’s law
   - Puzzle resolved

2. There is Spatial Sorting
   - Thick tails $\rightarrow$ bigger inequality in big cities
   - Extreme-skill compl.: Urban wage premium not due to skills
     $\rightarrow$ increasing over time $+$ urbanization $\uparrow \Rightarrow$ inequality $\uparrow$

3. Federal Income Taxation does affect local labor markets
   - Effect on location decisions: big cities are too small
   - Optimal level of taxation: progressive, but city-specific
ECONOMICS AND THE CITY

Jan Eeckhout†

†Barcelona GSE-UPF

Bojos per l’Economia
31 January, 2015
• Cities: dense, dirty, and polluted,...
Green Growth in Cities

- Cities: dense, dirty, and polluted,…
- Yet, green
Green Growth in Cities

- Cities: dense, dirty, and polluted,...
- Yet, green
- Large cities are more productive: urban wage premium = productivity premium
  Double city size and output grows by 4%
- But more expensive to live: elasticity wrt housing prices: 16%
- Large cities are more dense: more people in same space
  - Less consumption of energy
  - Less production of waste
Kleiber’s law
Kleiber (1947)

Fig. 1. Log. metabol. rate/log body weight
Kleiber’s law

- Energy consumption (metabolic rate) of animals and plants relates to their mass

\[ q \sim M^{3/4} \]

- Log-linear relationship
  - Cat 100 heavier than mouse, would use 31 times energy
  - For plants the exponent is close to 1

\( q \): metabolic rate; \( M \) body mass
From Biology to economics

- Energy efficiency: consumption of energy; production of waste
- But: mass is not size of the city, but economic productivity
- Economic productivity is correlated with size (Urban Wage Premium)
Urban Wage Premium
UK Data

Log wage bill

Log population

Coeff: 1.072 (.009), R²: .987, N: 178
Urban Energy Premium

14%

Log total energy vs Log wage bill

Coeff: .862 (.014), R2: .957, N: 178
### Urban Energy Premium
#### Breakdown By Source

**Table:** Energy Demand by Source

<table>
<thead>
<tr>
<th>Household</th>
<th>Transport</th>
<th>Industrial</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>33.9%</td>
<td>28.0%</td>
<td>38.1%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Urban Energy Premium
Breakdown By Source

Coefficient on wage bill

<table>
<thead>
<tr>
<th>Source</th>
<th>Total</th>
<th>Household</th>
<th>Transport</th>
<th>Industrial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>1.00</td>
<td>0.95</td>
<td>0.85</td>
<td>0.9</td>
</tr>
</tbody>
</table>
Urban Energy Premium

Why?

- Glaeser, Edward, *Triumph of the City*, 2011
- Energy Savings:
  1. Live in smaller space: less energy
  2. Apartments (vs. stand-alone buildings): more energy efficient
  3. Transportation: more efficient mass transportation (vs. car), walking, bike,...
Urban Waste Premium
10%

Coeff: .905 (.01), R2: .982, N: 141
## Urban Waste Premium
### Breakdown By Source

**Table:** Waste Supply by Source

<table>
<thead>
<tr>
<th></th>
<th>Household</th>
<th>Non-household</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recycled</td>
<td>35.1%</td>
<td>3.3%</td>
<td>38.4%</td>
</tr>
<tr>
<td>Non-recycled</td>
<td>54.1%</td>
<td>7.5%</td>
<td>61.6%</td>
</tr>
<tr>
<td>Total</td>
<td>89.2%</td>
<td>10.8%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Urban Waste Premium

Why?

- Housing: small space (no garages):
  - do not collect junk
  - buy less durables (furniture,...)
  - do not buy outdoors durables
A Policy Experiment
City-Specific Taxation

- From analysis on taxation results:
- Progressive taxation keeps workers from productive cities
- Productive cities are also clean

⇒ City-specific tax will:
   1. Increase population of big cities
   2. Increase productivity
   3. Shift people to cleaner living
A Policy Experiment
City-Specific Taxation

Aggregate growth: $-0.745\%$

% Growth in weighted energy usage

Log initial weighted energy usage

Aggregate growth: $-0.745\%$
A Policy Experiment

City-Specific Taxation

Aggregate growth: 

% Growth in weighted waste usage

Log initial weighted waste usage
ECONOMICS AND THE CITY

Jan Eeckhout†

†Barcelona GSE-UPF

Bojos per l’Economia
31 January, 2015