Essays on the Size Distribution of Cities

Hiroki Watanabe
Washington University in St. Louis

Follow this and additional works at: http://openscholarship.wustl.edu/art_sci_etds
Part of the Economics Commons

Recommended Citation
http://openscholarship.wustl.edu/art_sci_etds/426

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.
Washington University in St. Louis
Department of Economics

Dissertation Examination Committee:
Marcus Berliant, Chair
Sukkoo Kim
Alvin Murphy
John Nachbar
Bruce Petersen

Essays on the Size Distribution of Cities
by
Hiroki Watanabe

A dissertation presented to the Graduate School of Arts & Sciences of Washington University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

May 2015
St. Louis, Missouri
©2015, Hiroki Watanabe
TABLE OF CONTENTS

List of Figures ................................................................. v
List of Tables ................................................................. vi
Acknowledgements ........................................................... vii
Abstract .............................................................................. ix

1 Explaining the Size Distribution of Cities: X-treme Economies 1

1.1 Introduction and Motivation ............................................. 1

1.2 Modeling the Size Distribution of Cities 4
   1.2.1 A Model ............................................................... 4
   1.2.1.1 The Basic Model and Its Equilibrium ....................... 4
   1.2.1.2 Extensions of the Basic Model and Further Implications .. 9
   1.2.1.3 Stochastic Proportional Growth ............................ 12
   1.2.2 Empirical Implementation ....................................... 14

1.3 Relationship to the Literature 20
   1.3.1 The Older Literature ............................................... 20
   1.3.2 Recent Literature ................................................... 20
   1.3.3 Criticism of the Literature ....................................... 22
      1.3.3.1 How Insurance Reduces Population Movement ........ 22
      1.3.3.2 Possible Objections to the Criticism ...................... 25

1.4 Conclusions ................................................................. 28

2 A Scale-Free Transportation Network Explains City-Size Distribution 31

2.1 Introduction 31
   2.1.1 Cities on a Network ............................................... 33
   2.1.2 Some Transportation Networks Are Scale Free ................ 34
   2.1.3 The City-Size Distribution Is Scale Free Too ................ 36

2.2 Preliminaries 39
   2.2.1 ER Networks ....................................................... 39
   2.2.2 BA Networks ...................................................... 40

2.3 Model 40
   2.3.1 Location-Specific Commodities ............................... 41
   2.3.2 Network Infrastructure and Delivered Price .................. 42
2.3.3 Equilibrium ............................................. 43
2.3.4 How Does a Network Break Symmetry? ...................... 44
2.3.5 Transportation Cost Skews the City-Size Distribution .......... 46
2.3.6 Geodesic-Length Distribution. ............................. 47
2.3.7 City-Size Distribution .................................. 49
2.3.8 City-Size Distribution under Different Network Systems ........ 51

2.4 Empirical Implementation .................................. 53
2.4.1 Estimation Methods Employed .................................. 56
2.4.2 A Scale-Free Transportation Network Explains the City-Size Distribution ....... 56

2.5 Conclusion and Extensions .................................. 63

3 A Spatial Production Economy Explains Gross Metropolitan Product ........ 66

3.1 Introduction ................................................. 66

3.2 GMP Actualities ............................................ 71
3.2.1 Gibrat’s Law for GMP ....................................... 73
    3.2.1.1 Non-Parametric Estimation. ....................... 73
    3.2.1.2 Parametric Estimation. .............................. 74
3.2.2 Zipf’s Law for GMP ...................................... 75
3.2.3 City Size and GMP ........................................ 76

3.3 Model ....................................................... 79
3.3.1 Spatial Production Economy ................................. 79
3.3.2 Application ............................................... 84
    3.3.2.1 Intracity Equilibrium ............................... 84
    3.3.2.2 Intercity Equilibrium ............................... 88
3.3.3 Distribution of GMP ...................................... 91

3.4 Empirical Implementation .................................. 92
3.4.1 Distribution of GMP ...................................... 92
3.4.2 Scale Economies ......................................... 94

3.5 Conclusion and Extensions .................................. 95

A Appendix ..................................................... 97
A.1 A Model from the Literature Modified to Include Insurance .................. 97
    A.1.1 Notation ................................................. 97
    A.1.2 Insurance ............................................... 99
    Insurance when the initial state is the same for all cities ............. 101
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insurance starting when the state is an equilibrium at a given time</td>
<td>103</td>
</tr>
<tr>
<td>Extensions: Partial Insurance and Moving Cost</td>
<td>105</td>
</tr>
<tr>
<td>A.2 Positive Transport Costs</td>
<td>107</td>
</tr>
<tr>
<td>A.3 Proof of Proposition 2.3.1</td>
<td>113</td>
</tr>
<tr>
<td>A.4 Idea behind Geodesic Length (16)</td>
<td>113</td>
</tr>
<tr>
<td>A.5 Branching Factor</td>
<td>114</td>
</tr>
<tr>
<td>A.6 Maximum Spacing Estimation</td>
<td>115</td>
</tr>
<tr>
<td>A.7 Analysis of Variance</td>
<td>116</td>
</tr>
<tr>
<td>A.8 Unconditional Scale Economies.</td>
<td>117</td>
</tr>
<tr>
<td>A.9 The City-Size Distribution and Gibrat’s Law</td>
<td>118</td>
</tr>
</tbody>
</table>

References 119
# List of Figures

1. City-Size Distribution .................................................. 2
2. Empirical Density Function ............................................. 2
3. Estimated Distributions .................................................. 19
4. Completely Isolated Graph .............................................. 32
5. Complete Graph .......................................................... 32
6. ER Network ............................................................... 35
7. BA Network ............................................................... 35
8. Frequency Plot of the City-Size Distribution ......................... 37
9. Population Density by CBSA ............................................ 38
10. Accessibility and City Size .............................................. 44
11. $D(\tau)$ ................................................................. 46
12. Probability Density Function of Degree ............................... 52
13. Model Comparison (Belgium) ............................................ 58
14. Model Comparison (MA) .................................................. 58
15. Model Comparison (CBSA) .............................................. 59
16. Model Comparison (Places) ............................................. 59
17. MSE for Belgium and MA ............................................... 60
18. MSE for MSA and Places ............................................... 61
19. GMP per Capita .......................................................... 67
20. Population Density and GMP ............................................ 68
21. PDF Plot of GMP ......................................................... 68
22. Kernel Regression ....................................................... 74
23. OLS Regression ........................................................ 75
24. Zipf’s Law for Size and GMP .......................................... 76
25. OLS on GMP, GMP per Capita, and Housing over City Size .... 78
26. Commodity Flow ......................................................... 84
27. MLE on GMP and Housing ............................................. 93
28. Simulated City Size with Positive Transport Costs .................. 110
29. MLE with Transportation Costs .................................... 112
30. Kernel Density Estimation ............................................. 117
31. Kernel Density Estimation (Variance) ................................. 117
List of Tables

1. Summary Statistics ............................................................... 15
3. City Population Summary Statistics for Germany, Belgium, and the US .......... 24
4. Comparison of European and US Results ........................................ 25
5. Descriptive Statistics ................................................................ 55
6. Model Comparison ................................................................. 57
7. Descriptive Statistics ................................................................ 72
8. Ordinary Least Squared Estimate of Growth Rate .............................. 74
9. Rank-Size and Rank-GMP Regression ............................................ 75
10. OLS on GMP, GMP per Capita, and Housing over City Size .................. 79
11. Maximum Likelihood Estimation on GMP ....................................... 94
12. Positive Transport Costs ......................................................... 108
ACKNOWLEDGEMENTS

I have been privileged to meet incredible individuals throughout my doctoral program and I would not have come this far without them.

I was fortunate to study under Marcus Berliant. There is no uncle as insightful, patient, devoted or kind-hearted as he is. I am eternally indebted to him for his legendary tutelage over the years. I would also like to thank John Nachbar, Sukkoo Kim, Alvin Murphy and Kelly Bishop for their guidance throughout my doctoral studies.

My friends from the program, Jie Zheng, Atreyee Biswas, Luiggi Donayre, Aleksandr Yankelevich, Jeff Lee and Yin-Chi Wang have been always there for me and put me back on track when I got lost. I am very grateful that our paths crossed in St. Louis at that particular time in my life. They are all over the world today but I will always keep them close to my heart.

I would also like to thank Dorothy and Bruce Petersen and Beth Fisher for their practical guidance and pedagogical support.

Administrative support from the Economics Department and the Office for International Scholars and Students has been indispensable to lead my research as smoothly as possible. I also wish to acknowledge the John Stuart Mill Fellowship for the financial support.

Finally, I would like to thank Masahisa Fujita and Anne Conduit for their inspiring instruction and encouragement to pursue my academic career in a foreign country.
To my parents
Abstract of the Dissertation

Essays on the Size Distribution of Cities

by

Hiroki Watanabe

Doctor of Philosophy in Economics

Washington University in St. Louis, 2015

Professor Marcus Berliant, Chair

The dissertation is comprised of three essays that analyze the spatial distribution of people and economic activities from three distinct perspectives.

Chapter 1: Explaining the Size Distribution of Cities: X-treme Economies.1 The empirical regularity known as Zipf’s law or the rank-size rule has motivated development of a theoretical literature to explain it. We examine the assumptions on consumer behavior, particularly about their inability to insure against the city-level productivity shocks, implicitly used in this literature. With either self insurance or insurance markets, and either an arbitrarily small cost of moving or the assumption that consumers do not perfectly observe the shocks to firms’ technologies, the agents will never move. Even without these frictions, our analysis yields another equilibrium with insurance where consumers never move. Thus, insurance is a substitute for movement. We propose an alternative class of models, involving extreme risk against which consumers will not insure. Instead, they will move, generating a Fréchet distribution of city sizes that is empirically competitive with other models.

Chapter 2: A Scale-Free Network Structure Explains the City-Size Distribution.2 Zipf’s law is one of the best-known empirical regularities in urban economics. There is extensive

---

1Forthcoming at Quantitative Economics.
2Based on joint work with Professor Marcus Berliant.
research on the subject, where each city is treated symmetrically in terms of the cost of transactions with other cities. Recent developments in network theory facilitate the examination of an asymmetric transport network. In a scale-free network, the chance of observing extremes in network connections becomes higher than the Gaussian distribution predicts and therefore it explains the emergence of large clusters. The city-size distribution shares the same pattern. This paper decodes how accessibility of a city to other cities on the transportation network can boost its local economy and explains the city-size distribution as a result of its underlying transportation network structure.

**Chapter 3: A Spatial Production Economy Explains Gross Metropolitan Product.** It has long been known that the city-size distribution is fat tailed, drawing the interest of urban economists. In contrast, not much is known about the distribution of GDP at city level (henceforth referred to as gross metropolitan product, GMP). We build a model of the spatial economy that includes production and confirm the following empirical facts about the GMP counterpart of the city-size distribution. First, both Zipf’s and Gibrat’s law hold for the distribution of GMP as well. In particular the GMP distribution is well-traced by a lognormal distribution. Second, citywide aggregate production exhibits increasing returns to scale with respect to employment. In particular a 1% increase in employment leads to a 1.117% (or 1.180% in theory) increase in GMP. Agglomeration economies are explained as a result of an endogenous trade-off between externalities and land consumption of consumers.
1.1 Introduction and Motivation

A small industry has developed that seeks to provide a theory to explain a singular but robust stylized fact in urban growth: the size distribution of cities. Zipf’s law or the rank-size rule, as applied to the size distribution of cities, states that for any country, the rank of a city according to population (for example, New York is ranked number one in the US) multiplied by its population is constant. Thus, Los Angeles has half the population of New York, whereas Chicago has one third the population of New York. This stylized fact holds across many countries and time periods (see Soo [Soo05]), but it is only one fact. In general, it is connected to Gibrat’s law, stating that stochastic proportional growth tends to a lognormal distribution. The most compelling empirical work in this area shows that the size distribution of cities is lognormal (Eeckhout [Eec04]) when the data is not cut off at an arbitrary rank or population. For those unfamiliar with the empirics associated with this literature, we display in figure 1(a) a graph of Eeckhout’s data, consisting of more than 25,000 places from U.S. Census 2000.

Since population on the horizontal axis and rank on the vertical axis are both plotted in log scales, the rank-size rule, taken literally, would say that the plot should be linear with slope −1. Deviations from the rule or law at the top and bottom of the size distribution are documented and discussed in the literature. See Gabaix and Ioannides [GI04] for a fine survey of the entire area of research.

Further orientation with the data will prove useful so that the finer details of the distribution might be seen. The log-log plot is rather uninformative since very different distributions can
appear similar because the majority of observations are bunched in the middle where there is little variation in the log-log scale. To that end, in figure 1(b) we provide a graph of the empirical distribution function, whereas in figure 2 we provide the density function.

Explanation of the stylized fact illustrated in these figures by a theory has long been an objective of urban economists; it is quite robust, but also very difficult to theorize about. Three recent articles, Eeckhout [Eec04], Duranton [Dur07], and Rossi-Hansberg and Wright [RHW07], have tackled this issue head on. The general methodology in this literature is as follows. A city is defined as a set of firms that receive a common technological shock to their production functions. Generally speaking, the shock is observed each period before the agents make their decisions. Consumers are freely mobile between cities. A model of de-
mand and supply is formulated, generally relying on specific functional forms to obtain an analytical solution for equilibrium prices and quantities as a function of shock realizations. The key equation obtained from the models is the reduced form for the evolution of city population over time. Frequently (but not universally) this equation yields stochastic proportional growth for each city’s population, where the stochastic component is derived from the city-specific technology shock. Then Gibrat’s law is applied. The lognormal distribution matches Zipf’s law well for the upper portion of the distribution.

The contribution of our work is as follows. First, we propose a new stochastic model of technological innovation in cities under perfect competition, giving rise in the limit to a generalized extreme value distribution of city sizes in aggregate, where the Fisher-Tippett Theorem replaces the central limit theorem and Gibrat’s law in a natural way. This model and its implications are robust against the introduction of self-insurance or insurance into the framework. The other models are generally not robust to the introduction of self-insurance or insurance, as we illustrate formally for one example from the literature in appendix A.1. Our model is empirically competitive with other models of the size distribution of cities. In particular, the error in the estimate is very close to Eeckhout’s [Ee04] for the lognormal distribution in the data on places, but better than Eeckhout’s [Ee04] for the MSA data.

All of these models, including ours, feature uncertainty that affects consumers through the budget constraint only. By self-insurance, we mean that consumers have an integrated budget constraint over time and know the distribution of future realizations of the random variables. Thus, they can smooth consumption. Since the consumers are risk averse, insurance or self-insurance is a substitute for migration. In the theoretical models, since moving is a discrete choice, partial insurance is never chosen in equilibrium by an individual agent (see the last subsection of appendix A.1).

The paper is organized as follows. First, in section 1.2, we propose a new type of model to explain the size distribution of cities, and implement it empirically. Only in section 1.3 shall we discuss in detail the related literature that attempts to refine the stylized fact, namely the rank-size
rule, and explain it. Then we shall raise specific objections, involving insurance or self-insurance against city-level risk, to these models. Section 1.4 discusses our conclusions and directions for future work. In appendix A.1 we introduce Eeckhout’s [Eco4] model and modify it to make the objections raised in section 1.3 formal for a specific example, whereas in appendix A.2 we examine our model with positive transport costs.

1.2 Modeling the Size Distribution of Cities

1.2.1 A Model

1.2.1.1 The Basic Model and Its Equilibrium

This model is loosely based on Duranton [Duro7], but in the context of perfect competition instead of monopolistic competition. It can also be viewed as a slice of a larger model that would include both our model and the model of Eaton and Kortum [EKo2]. Our model adds labor and consumer mobility, whereas their model has them locationally fixed. In contrast with the other models in the literature, there is economy-wide risk in addition to city-level risk. But this in itself is not sufficient to generate consumer movement. For example, if all cities faced correlated shocks at each time, consumers could still insure against this risk by smoothing their consumption through borrowing and saving. Thus, we employ a more extreme form of aggregate risk.

Time is discrete and all consumers are infinitely lived. Assume that there are many cities (indexed by \( i = 1, \ldots, m \)) and many industries, each producing one consumption commodity (indexed by \( j = 1, \ldots, n \)). All commodities are freely mobile. The production function for commodity \( j \) in city \( i \) at time \( t \) is given by

\[
y_{ijt} = A_{ijt} \cdot l_{ijt}
\]

where \( y_{ijt} \) is the output of commodity \( j \) in city \( i \) at time \( t \), and \( l_{ijt} \) is labor input.\(^1\) The random

\(^1\)The assumption of Starrett’s [Sta78] spatial impossibility theorem that is violated by this model is the assumption of location-independent production sets.
variable $A_{ijt} \in \mathbb{R}_{++}$ will be discussed in detail shortly. Suppose that each consumer supplies 1 unit of labor inelastically and that the total number of consumers as well as total labor supply is given by $N$. We justify the assumption of perfect competition by implicitly assuming that there is a large number of firms in each city capable of producing a commodity using a constant returns technology, but all experiencing the same city-wide technology shock.

In each time period $t$, each city $i$ receives a random draw for its productivity in producing commodity $j$, namely $A_{ijt}$. Since we will be using the Fisher-Tippett limit theorem from extreme value theory rather than the central limit theorem, there is no requirement that these random variables be independent. It is assumed that with probability 1, the random draws for 2 industries at time $t$ for city $i$ are not both maximal among all cities for these given industries. In equilibrium, only the cities with the highest draw of the random variable for some industry will have employees and population. (Alternatively, we could simply classify cities exogenously by industry, and assume that a city in an industry receives only a draw for that industry.) Extensions that imply several cities produce in equilibrium will be discussed shortly, but first we must explain the basic model.

The wage rate for the (freely mobile) population of consumers is given by $w(t)$. In equilibrium, it will be the same across industries.

As is standard in this literature, the utility function of a consumer at time $t$ is given by

$$u(t) = \sum_{j=1}^{n} \frac{1}{n} c_j(t)^\gamma$$

where $c_j(t)$ is the consumption of commodity $j$ by a consumer at time $t$ and $\gamma \in (0, 1)$. Let $p_j(t)$ be the price of commodity $j$ at time $t$. Assuming that commodities are freely transportable, a consumer’s budget constraint at time $t$ is

$$\sum_{j=1}^{n} p_j(t) \cdot c_j(t) = w(t)$$
Let $\lambda(t)$ be the Lagrange multiplier associated with the budget constraint in the consumer optimization problem. Standard calculations yield demand for commodity $j$ at time $t$ for a single consumer $d_j(t)$:

$$d_j(t) = \left( \frac{\gamma'}{-\lambda(t) \cdot n \cdot p_j(t)} \right)^{\frac{1}{1-\gamma}}$$

Aggregate demand is given by

$$N \cdot d_j(t) = N \left( \frac{\gamma'}{-\lambda(t) \cdot n \cdot p_j(t)} \right)^{\frac{1}{1-\gamma}}$$

To reduce notation, for $j = 1, ..., n$, define $i^*$ to be the city with $A_{i^*jt} = \max_{1 \leq i \leq m, 0 \leq t' \leq t} A_{ijt'}$. Profit optimization yields, for each $t$:

$$p_j(t) \cdot A_{i^*jt} = w(t)$$

Here we are assuming total recall, in that the best technology from the past is remembered, so new technologies are not used unless they are better than all the old ones. Also, only the best technology in industry $j$ survives, where the best is across all cities and previous time periods. This assumption is made for convenience. We discuss it more below.

Hence

$$p_j(t) = \frac{w(t)}{A_{i^*jt}} \quad (1)$$

In other words, even though wage is constant across occupied cities, output price varies inversely with the production shock. Consumption commodity market clearance requires, for each $t$:

$$l_{i^*jt} \cdot A_{i^*jt} = N \cdot d_j(t) = N \left( \frac{\gamma'}{-\lambda(t) \cdot n \cdot p_j(t)} \right)^{\frac{1}{1-\gamma}} \quad (2)$$

This is the key equation for our analysis.
Labor market clearance requires, for each $t$:

$$\sum_{j=1}^{n} l_{rjt} = N \quad (3)$$

Setting the constant $\kappa(t)$ to be

$$\kappa(t) = N \left( \frac{\gamma}{-\lambda(t) \cdot n \cdot w(t)} \right)^{\frac{1}{1-\gamma}}$$

and using (1) and (2), we obtain

$$l_{rjt} \cdot (A_{rjt})^{\frac{\gamma}{1-\gamma}} = \kappa(t)$$

Hence

$$l_{rjt} = \kappa(t) \cdot (A_{rjt})^{\frac{\gamma}{1-\gamma}} \quad (4)$$

Since $\gamma < 1$, labor usage $l_{rjt}$ and the shock $A_{rjt}$ are positively correlated. Notice that cities that do not have an industry with the largest shock in that industry at time $t$ are empty.

Existence of an equilibrium is not an issue here, since the equilibrium prices and quantities can be solved analytically. For example, at $t = 1$, setting $p_{1}(1) = 1$, then $w(1) = A_{r11}$, $p_{j}(1) = A_{r1j}/A_{rjt}$, $\lambda(1) = -\gamma n A_{r1j}^{\frac{\gamma}{1-\gamma}}$, $l_{rj1} = N \left( \frac{\gamma}{-\lambda(1)n A_{r1j}} \right)^{\frac{1}{1-\gamma}} A_{rj1}^{\frac{\gamma}{1-\gamma}}$, and so forth. Thus, equilibrium is also unique.

The original work on the asymptotic distribution of maxima drawn from a distribution is due to Fisher and Tippett [FT28]. Modern, more general treatments are given in Coles [Col01] and Embrechts, Kluppelberg, and Mikosch [EKM97]. We shall return to a discussion of extreme value theory momentarily, but first we will draw the implications for our analysis.

The bottom line from this literature is that $A_{rjt}$ has an asymptotic distribution of the following
form, known as the generalized extreme value (GEV) distribution:

\[
F_{GEV}(x) = \begin{cases} \exp \left\{ - \left[ 1 + \xi \cdot \left( \frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} & \text{when } \xi \neq 0 \\ \exp \left\{ - \exp \left[ - \left( \frac{x - \mu}{\sigma} \right) \right] \right\} & \text{when } \xi = 0 \end{cases}
\]

Notice that there are 3 free parameters to be estimated here, namely \(\mu\), \(\sigma\), and \(\xi\). Also notice that to use rank as the left hand side variable in the regression, one simply computes \(1 - F_{GEV}(x)\). But from a pragmatic point of view, it is easier to use \(\ln(F_{GEV}(x))\) as the left hand side variable.

If there are no upper or lower bounds on the distribution, then \(\xi = 0\) and the distribution is Gumbel. If there is an upper bound on the distribution, then \(\xi < 0\) and the distribution is reverse Weibull. If there is a lower bound on the distribution, for example 0 in our case, then \(\xi > 0\) and the distribution is Fréchet.

Substituting (4),

\[
\ln(F(l)) = \begin{cases} - \left[ 1 + \xi \cdot \left( \frac{1}{\kappa(t)} \right)^{1-\gamma} - \mu \right] / \sigma & \text{when } \xi \neq 0 \\ - \exp \left[ - \left( \frac{1}{\kappa(t)} \right)^{1-\gamma} - \mu \right] / \sigma & \text{when } \xi = 0 \end{cases}
\]

Notice that if we use cross section data, then \(t\) and hence \(\kappa(t)\) is constant. Thus, in addition to the 3 standard parameters for the GEV distribution of \(A_{rjt}\) (namely \(\mu\), \(\sigma\), and \(\xi\)), for the distribution of \(l_{rjt}\) there are two additional parameters, namely \(\kappa\) and \(\gamma\), that arise from our economic model.

Now that the basic model is fully developed, we can discuss why, unlike other models in this literature, consumers will not want to insure against this risk. Instead, they will move. If only a small percentage of cities produce at any time, then insurance would cost only slightly less than the wage, so the consumers might as well move and receive the wage in each period. For example, to keep things simple suppose that there are 100 industries (or consumption commodities) and 100 cities in each industry (that is, each city is capable of producing only one commodity). Then there is only one city producing in each industry at each given time, and 100 cities out of 10,000
producing in each given time. As time plays out, as long as some consumers are willing to move, each of the cities producing at a given time will eventually be replaced by another in the industry. The city using old technology has zero wage and no production. So if some workers don’t move, their average wage tends to one percent of the expected new wage with time. Under symmetry of cities in an industry, actuarially fair insurance would cost 99% of the expected new wage. In other words, if workers move they will receive the wage next period, but if they insure they will receive 1% of the wage next period. The only way workers won’t move is if they all agree to use old, frozen technology in each industry, and collude so that none will move for a higher wage. In contrast, we assume competitive behavior.

This is the main idea motivating our specific model. Next, we discuss extensions and the intuition behind why this idea is robust.

1.2.1.2 Extensions of the Basic Model and Further Implications

In fact, what we have presented is an extreme example. All that is needed to induce consumers to reject insurance and move is that the probability of unemployment next period is greater than zero if they don’t move. To obtain stronger results, for example the GEV distribution, stronger assumptions are required. Thus, there are many models like this in which consumers will not take up insurance, but that do not require such strong assumptions. We provide a simple one that is tractable.

We claim that the choice of insurance or moving is essentially a bang-bang phenomenon, not only in this model but in other models of stochastic growth belonging to the literature that will be surveyed in section 1.3. That is, generically one or the other will be better for consumers, so in equilibrium they will not coexist. Moreover, in equilibrium there will be no partial insurance. To see this, notice first that utilities are not state-dependent, so the state only directly affects budget constraints. Second, the decision to move is a discrete one: Either all of a moving cost or none of it is incurred by a particular consumer. If competition forces insurance to be priced competitively, implying both that consumer cost is proportional to price and that it is actuarially
fair, then risk averse consumers will always want to fully insure or move, facing no uncertainty in equilibrium. The consumers must consider whether the moving cost or the cost of full insurance is cheaper. Generically these exogenous parameters are unequal, so only one or the other will be observed in equilibrium. Partial insurance will not result unless there is some defect in insurance markets; but the random shock in this entire class of models is assumed to be observed by all agents. Generically, none would predict that consumers use partial insurance. We prove this more formally in the appendix A.1 in the case where moving is costly and insurance is actuarially fair.

Given the structure of the model and the others in the literature, it is much more natural to introduce a market imperfection in the labor market: labor heterogeneity and adverse selection, moral hazard, or search frictions, for example. This new source of uncertainty or asymmetric information requires an additional dimension for states of nature beyond the states we have specified for production shocks. It leads to a different form of a distortion or market imperfection than the one in this literature, since for example labor supply might be distorted. Although individual labor supply is inelastic in our basic model, it is elastic for example in Eeckhout [Eec04]; see appendix A.1 below. Elastic labor supply could easily be put into our model in an additively separable way at the cost of further notation. The consequences of a distortion in the labor market would be very different from the introduction of an exogenous mobility cost that varies between zero and infinity, as described in the previous paragraph. Full or partial insurance would have to be defined over states of the world associated with a new source of uncertainty or asymmetric information related to the labor market, in contrast with the one already in the model that is related to production shocks.

Returning to our basic model, the consumers still might want to insure against aggregate wage volatility (namely movement in \( w(t) \) over time) by saving and borrowing to smooth consumption, but their spatial distribution is still as we have laid out.

Returning now to our assumptions and extreme value theory, the original theory of Fisher and Tippett presumed that, fixing \( j \), the random variables, \( A_{ij} \) in our case, were i.i.d. across \( i \)
of course, in our context this makes little sense. In general, the city with the best technology for some good \( j \) at a particular time \( t \) is more likely to innovate and produce a better technology for the next period than an arbitrary city. Moreover, it is possible that cities nearby are more likely to innovate than an arbitrary city. Fortunately, much progress has been made in extreme value theory since 1928. The modern versions of the Fisher-Tippett theorem, as given by Coles ([Col01], Theorem 5.1) and Embrechts, Kluppelberg, and Mikosch ([EKM97], Theorem 4.4.1) allow some dependence. Specifically, what is required is that the sequence of random variables be stationary and that a form of asymptotic independence (as blocks of random variables become farther apart in time) hold. Since temporal (as well as spatial) correlation is allowed, the model can explain the persistence of an industry in a given city over time. For example, the assumption that the process is stationary imposes some symmetry on the spatial correlation, in that the influence of neighbors on the productivity of one reference location is the same, independent of the reference location. However, we note that even the modern versions of the Fisher-Tippett theorem we have cited give only sufficient conditions for convergence to the GEV distribution. There are yet further generalizations to non-stationary processes; see Coles ([Col01], chapter 6) for example. So asymmetries in space, implying that the process is not stationary, can still lead to the GEV distribution.

Returning to the case of i.i.d. technology draws, an implication of extreme value theory (Embrechts, Kluppelberg, and Mikosch ([EKM97], chapter 5.4)) is that the time between new record draws of technology in an industry grow in a roughly exponential fashion with the passage of time. This implication of the theory might not hold in more general settings, for example non-stationary ones.

It is also important to note that the model and results can be extended to the case where more than one city in an industry produces. This could happen, for example, if there is transportation cost for consumption goods between cities, so a city with a high realization of productivity for a commodity, but not the highest, might serve a local market. It turns out that extreme value theory applies not only to the maximum of a sequence of random variables, but also to the upper

---

Footnote:

2 An easy way to fit our structure into the theory is to fix an industry \( j \) and imagine that at each time \( t \), there are \( m \) subperiods. A city \( i \) draws its random variable \( A_{ijt} \) in subperiod \( i \) of time \( t \).
order statistics. A detailed discussion of the results can be found in Embrechts, Kluppelberg, and Mikosch ([EKM97], Section 4.2). These extensions of the model require a simulation approach, as the analytics are difficult. Specifically, the calculation of aggregate demand on the right hand side of equation (2) becomes difficult due to the endogeneity of market area. Our simulations appear in appendix A.2 below.

A couple more remarks are in order. First, the role of having different industries \( j \), as in the other models in the literature, is to generate a full distribution of limiting populations rather than just one realization of the asymptotic distribution of city populations. Second, in contrast with other models in the literature, the cities without the best technology for some industry at a given time have zero population, so they don’t show up in the data because they are rural.

1.2.1.3 Stochastic Proportional Growth

As a complement to our basic analysis of the model, it is interesting to see under what conditions our model will generate stochastic proportional growth in (occupied) city populations. To examine this, we must specialize and reinterpret slightly the stochastic part of our model, inspired by Eeckhout ([Eco04], p. 1447). Suppose that a primitive productivity random variable is generated by the AR(1) process:

\[
B_{ijt} = k \cdot B_{ij(t-1)} + \frac{1 - \gamma}{\gamma} \cdot \epsilon_{ij(t-1)}
\]

(6)

where \( \epsilon_{ij(t-1)} \) is i.i.d. with mean 0 and finite variance, and where \( 0 < k < 1 \). For the purpose of approximation, we will be taking \( k \) close to 1. Then define the reduced form random variable \( A_{ijt} \) by:

\[
A_{ijt} \equiv \exp(B_{ijt})
\]

Our previous analysis applies to this more specific model of \( A_{ijt} \), for instance the aforementioned Theorem 4.4.1 of Embrechts, Kluppelberg, and Mikosch [EKM97], along with all of the results in the subsections above. But with this additional structure, we can say more.
Consistent with our notation:

For \( j = 1, \ldots, n \), let \( \tilde{r} \) be such that \( B_{\tilde{r},jt} = \max_{1 \leq i \leq m, 0 \leq t' \leq t} B_{ijt'} \)

If the \( \varepsilon_{ijt} \) are small, we claim that

\[
B_{\tilde{r},jt} \approx k \cdot B_{\tilde{r},j(t-1)} + \frac{1 - \gamma'}{\gamma'} \cdot \varepsilon_{\tilde{r},j(t-1)}
\]

in the sense that the distributions of the two sides of this expression viewed at time \( t - 1 \) are close.

The reasoning behind this approximation is as follows. Fix industry \( j \). If \( B_{\tilde{r},j(t-1)} \gg k \cdot B_{\tilde{r},j(t-1)} \) for all \( 1 \leq i \leq m, 0 \leq t' \leq t - 1, \tilde{r}' \neq \tilde{r} \), then the city with the maximal draw remains the same between periods \( t - 1 \) and \( t \), so the approximation holds according to equation (6). If \( B_{\tilde{r},j(t-1)} \approx k \cdot B_{\tilde{r},j(t-1)} \) for some \( 1 \leq i' \leq m, 0 \leq t' \leq t - 1, \tilde{r}' \neq \tilde{r} \), then the distribution of \( k \cdot B_{\tilde{r},j(t-1)} + \frac{1 - \gamma'}{\gamma'} \cdot \varepsilon_{\tilde{r},j(t-1)} \) conditional on \( B_{\tilde{r},j(t-1)} \) is close to the distribution of \( k \cdot B_{\tilde{r},j(t-1)} + \frac{1 - \gamma'}{\gamma'} \cdot \varepsilon_{\tilde{r}',j(t-1)} \) conditional on \( B_{\tilde{r}',j(t-1)} \), so the approximation holds.

Dividing equation (4) at time \( t \) by its value at time \( t - 1 \) to the power \( k \),

\[
\frac{l_{\tilde{r},jt}}{[l_{\tilde{r},j(t-1)}]^k} = \frac{\kappa(t)}{[\kappa(t-1)]^k} \cdot \left( \frac{A_{\tilde{r},jt}}{[A_{\tilde{r},j(t-1)}]^k} \right)^{\frac{\gamma}{1 - \gamma}} \quad \text{for each industry } j = 1, \ldots, n \tag{7}
\]

Using (4) and (3),

\[
N = \sum_{j=1}^{n} l_{\tilde{r},jt} = \kappa(t) \cdot \sum_{j=1}^{n} (A_{\tilde{r},jt})^{\frac{\gamma}{1 - \gamma}}
\]

Now

\[
\lim_{n \to \infty} \frac{\sum_{j=1}^{n} (A_{\tilde{r},jt})^{\frac{\gamma}{1 - \gamma}}}{n} = E[(A_{\tilde{r},jt})^{\frac{\gamma}{1 - \gamma}}]
\]

13
Hence for $k$ close to 1,
\[
\frac{\kappa(t)}{[\kappa(t-1)]^k} \approx \frac{E[(A_{i,j(t-1)})^{\gamma}]}{E[(A_{i,j})^{\gamma}]} \approx 1
\]

Taking logarithms of both sides of equation (7):
\[
\ln(l_{i,j(t)}) = k \cdot \ln(l_{i,j(t-1)}) + \frac{\gamma}{1-\gamma} \cdot \frac{1-\gamma}{\gamma} \cdot \varepsilon_{i,j(t-1)}
\]
\[
= k \cdot \ln(l_{i,j(t-1)}) + \varepsilon_{i,j(t-1)}
\]
\[
\approx \ln(l_{i,j(t-1)}) + \varepsilon_{i,j(t-1)}
\]

This last equation is the form of stochastic proportional city population growth obtained in Eeckhout [Eeco4].

The assumption that $k < 1$ is essential, in the sense that $k = 1$ yields Gibrat’s law and a lognormal distribution for occupied cities. The assumption $k < 1$ implies the asymptotic independence used for modern variants of the Fisher-Tippett theorem. In contrast, $k = 1$ implies some permanent path dependence. Another way of framing the arguments in this subsection is that the order of limits in $k$ and $t$ matters.

\subsection*{1.2.2 Empirical Implementation}

Notice that we are not overly concerned with identification of the 5 parameters in equation (5). In essence, the parameters are identified by the functional form itself. The economic interpretation of these variables is as follows. The three parameters of the GEV distribution, $\mu$, $\sigma$, and $\xi$, are analogous to the mean and variance of the lognormal distribution estimated by Eeckhout, or the regression coefficients estimated for Zipf’s law using a log-log regression. They have no direct economic interpretation. Since $\gamma$ and $\kappa$ are derived from the model, they do have an economic interpretation. Standard calculations tell us that $\frac{1}{1-\gamma}$ is the elasticity of substitution for consumers between consumption commodities. The endogenous variable $\kappa$ is more difficult to interpret, since it involves a number of endogenous variables as well as random variables.
But equation (4) gives us the equilibrium relationship between the random variable representing productivity in an industry (exogenous) and employment in that industry (endogenous). So $\kappa(t)$ tells us equilibrium employment in an industry where one unit of labor produces one unit of consumption commodity.

We use the Census 2000 data set also used by Eeckhout. Table 1 gives the summary statistics for this data along with the MSA-level data that we use later for comparison.\(^3\)

<table>
<thead>
<tr>
<th>Unit</th>
<th>Sample Size</th>
<th>Mean</th>
<th>Variance</th>
<th>Median</th>
<th>Mode</th>
<th>Max</th>
<th>Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Place</td>
<td>25,358</td>
<td>8.232E+03</td>
<td>4.677E+09</td>
<td>1,338</td>
<td>86</td>
<td>8,008,278</td>
<td>1</td>
</tr>
<tr>
<td>MSA</td>
<td>922</td>
<td>2.837E+05</td>
<td>9.490E+11</td>
<td>71,800.5</td>
<td>20,411</td>
<td>18,323,002</td>
<td>13,004</td>
</tr>
</tbody>
</table>

Table 1. Summary Statistics for US data

As noted in the sources we cite for extreme value theory, the most common method of estimating extreme value distributions is to use maximum likelihood. The maximum likelihood estimator (MLE) does not yield the smallest Kolmogorov-Smirnov (KS) statistic in our data set. The KS statistic measures the maximum distance between a sample distribution and its estimate. As noted by Goldstein, Morris, and Yen [GMY04] in the context of social networks and later by Eeckhout (2009) in the context of the size distribution of cities, using a simple log-log regression can lead to serious statistical problems. The use of MLE and the KS statistic is preferred. It is interesting to note that both the literature on estimation of the GEV distribution and the literature on Zipf’s law seem to be (independently) converging on MLE as the preferred method of estimation.

For purposes of comparison with Eeckhout [Eec04], we produce estimates using each of the lognormal (his) distribution and the generalized extreme value (our) distribution using equation (5), for both maximum likelihood estimation and minimization of the KS statistic (MinKS). We also report an estimate using the double Pareto lognormal distribution (DPL) from Giesen, Zimmermann, and Suedekum [GZS10] for comparison. Table 2 below summarizes the estima-

\(^3\)For a definition of the spatial units used by the Census, see for example [http://www.genesys-sampling.com/pages/Template2/site2/61/default.aspx](http://www.genesys-sampling.com/pages/Template2/site2/61/default.aspx)
The results of maximum likelihood estimation for the lognormal distribution are identical to Eeckhout’s. The rightmost columns contain the KS statistic, the log likelihood of the estimates (LogLH), the Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC).
<table>
<thead>
<tr>
<th>Unit</th>
<th>Distribution</th>
<th>Method</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$\xi$</th>
<th>$\kappa$</th>
<th>$\gamma$</th>
<th>KS</th>
<th>LogLH</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Place</td>
<td>Lognormal</td>
<td>MLE</td>
<td>7.278</td>
<td>1.754</td>
<td></td>
<td></td>
<td></td>
<td>1.895E-02</td>
<td>-2.347E+05</td>
<td>4.695E+05</td>
<td>4.6957E+05</td>
</tr>
<tr>
<td></td>
<td>GEV</td>
<td>MLE</td>
<td>1.410</td>
<td>.3096</td>
<td>-2.902E-02</td>
<td>57.15</td>
<td>.8827</td>
<td>8.638E-03</td>
<td>-2.3467E+05</td>
<td>4.6935E+05</td>
<td>4.6939E+05</td>
</tr>
<tr>
<td></td>
<td>Lognormal</td>
<td>MinKS</td>
<td>7.249</td>
<td>1.738</td>
<td></td>
<td></td>
<td></td>
<td>1.336E-02</td>
<td>-2.3478E+05</td>
<td>4.6956E+05</td>
<td>4.6958E+05</td>
</tr>
<tr>
<td></td>
<td>GEV</td>
<td>MinKS</td>
<td>1.592</td>
<td>.6127</td>
<td>1.592</td>
<td>102.9</td>
<td>.8100</td>
<td>6.970E-03</td>
<td>-2.3470E+05</td>
<td>4.6941E+05</td>
<td>4.6945E+05</td>
</tr>
<tr>
<td>MSA</td>
<td>Lognormal</td>
<td>MLE</td>
<td>11.46</td>
<td>1.190</td>
<td></td>
<td></td>
<td></td>
<td>9.426E-02</td>
<td>-1.203E+04</td>
<td>2.406E+04</td>
<td>2.407E+04</td>
</tr>
<tr>
<td></td>
<td>GEV</td>
<td>MLE</td>
<td>4.295</td>
<td>2.192</td>
<td>.6383</td>
<td>4552</td>
<td>.6276</td>
<td>2.582E-02</td>
<td>-1.190E+04</td>
<td>2.382E+04</td>
<td>2.384E+04</td>
</tr>
<tr>
<td>DPL</td>
<td>MLE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.694E+05</td>
<td>4.695E+05</td>
<td>4.695E+05</td>
</tr>
<tr>
<td>Germany</td>
<td>Lognormal</td>
<td>MLE</td>
<td>11.74</td>
<td>.9865</td>
<td></td>
<td></td>
<td></td>
<td>.1737</td>
<td>-2577</td>
<td>5158</td>
<td>5164</td>
</tr>
<tr>
<td>(1998)</td>
<td>GEV</td>
<td>MLE</td>
<td>2.242</td>
<td>1.159</td>
<td>1.081</td>
<td>3.179E+04</td>
<td>.5085</td>
<td>.05415</td>
<td>-2509</td>
<td>5027</td>
<td>5044</td>
</tr>
<tr>
<td>Belgium</td>
<td>Lognormal</td>
<td>MLE</td>
<td>1.84</td>
<td>.5697</td>
<td></td>
<td></td>
<td></td>
<td>.2030</td>
<td>-806.3</td>
<td>1617</td>
<td>1594</td>
</tr>
<tr>
<td>(2000)</td>
<td>GEV</td>
<td>MLE</td>
<td>1.529</td>
<td>.01775</td>
<td>.3513</td>
<td>.6682</td>
<td>.9626</td>
<td>.1064</td>
<td>-786.6</td>
<td>1583</td>
<td>1621</td>
</tr>
</tbody>
</table>

Table 2. Parameter Estimates and Related Statistics - US and Europe
In the interest of full disclosure, we report both the MLE and MinKS estimates in table 2. Notice that the MLE estimate implies a reverse Weibull distribution whereas MinKS estimates imply a Fréchet distribution. Since city sizes do not fall below zero, we expect the distribution to follow a Fréchet distribution. MLE predicts otherwise due to the large, uncensored data set containing places. The estimated Fréchet distribution under MinKS implies that the smallest place will have population 1.582, and two places actually fall below this size. Indeed, once we truncate the data to MSA’s, MLE predicts a Fréchet distribution. So the reverse Weibull GEV distribution is driven by extremely small populations in the sample of places.

Of course, the comparison between lognormal and GEV is not quite fair. In general, the more parameters a distribution has, the better its fit to data. There are only two parameters in the lognormal distribution whereas there are five parameters in our distribution, and these parameters do not contain the parameters used for the lognormal distribution. In Table 2, we report the Akaike and Bayesian information criteria, that penalize distributions with more parameters. Smaller values for these criteria mean better performance. Those two statistics indicate that the lognormal and our distribution are still comparable when an adjustment for the number of parameters is made. In particular for the uncensored place data, after penalizing each estimate for the number of parameters used, where the penalty is larger for the GEV estimate, the error is quite similar, with GEV slightly ahead. The penalties are actually quite small relative to the log likelihood, since the data sets are so large. There is more divergence between lognormal and GEV in the error for the MSA estimates. Clearly, as Eeckhout (2004) points out, there are problems with truncation of this data. On the other hand, it seems quite odd to give places with just a few people in them the same weight as, say, New York City in the data. Implicitly in the place data, all places have weight 1. In the MSA data, places above a certain population have weight 1, whereas all other weights are set to 0. There is likely a way to weight the data better than these extreme cases, but we do not attempt a formal theory of data weighting for these estimates.4 So

4As Eeckhout ([Eeco4], pp. 1434-1436) points out, the theoretical definition of a city as those firms that experience a common shock (perhaps assuming spatial independence of shocks) should drive the empirical unit used. In places with only a few inhabitants, it is difficult to see how to apply this definition. On the empirical side, Eeckhout (2004, Figure 7) finds some anomalies at the low end of the distribution. In our opinion, random models of city growth
we do not completely discount the MSA estimates (as Eeckhout might), but rather await a less extreme data weighting scheme than the two standard ones. The truth probably lies somewhere in between. This appears to be an interesting topic for future research.

Graphically (in color), the estimates and data plots follow in figure 3.

![Graph](image)

**(a) MLE of lognormal distribution**

**Figure 3.**

In summary, estimates using the generalized extreme value distribution are quite competitive.

We ran simulations of our model with positive transportation cost. The results and discussion can be found in appendix A.2. Related to this, Hsu, Mori, and Smith [WTTT14] study a random might not be appropriate for small places.
growth model of the city size distribution when city connections matter.

1.3 RELATIONSHIP TO THE LITERATURE

1.3.1 THE OLDER LITERATURE

The innovative work of Gabaix [Gab99a, Gab99b] is the source from which the modern literature on the size distribution of cities flows. This work uses an overlapping generations structure where consumers live for two periods. It is assumed that moving costs are so high that consumers can only choose their location (city) when they are young. This location decision is made after shocks to production and amenities are realized for that period, and known to all. The consumer/workers cannot move again when old. The wages or income for the old in a city are never even specified, and it is simply assumed that the young make their decisions in a myopic manner. Moreover, the availability of insurance or capital markets is never discussed, so it is unknown whether the young can hedge against uncertainty about their wage when they are old in the city they choose.

If the old people are immobile, why is this important? It is important because when the young make their decisions, they can anticipate what happens when they are old, and might change their minds about their location decisions when young. In other words, they won’t behave myopically. Without myopia, insurance becomes important.

1.3.2 RECENT LITERATURE

Chief among recent work are Rossi-Hansberg and Wright [RHW07], Duranton [Dur06], Eeckhout [Eec04] and Duranton [Dur07]. We focus on the latter two.

Eeckhout’s model has consumers who are infinitely lived with foresight and who can move each period. There are technological shocks to production in each city in each time period. It is movement of the consumer/worker population in response to these shocks that generates
Gibrat’s law. The shocks generate changes in equilibrium wages, rents, and congestion across time and space that correspond to the consumer movements that equalize utility levels across space at each time. On p. 1445, the following statement is made: “Moreover, because there is no aggregate uncertainty over different locations, and because capital markets are perfect, the location decision in each period depends only on the current period utility. The problem is therefore a static problem of maximizing current utility for a given population distribution, and the population distribution must be such that in all cities, the population $S_{it}$ equates utilities across cities.”

Here we wish to make an important distinction between transfers of consumption across time, namely perfect capital markets, and across states, namely complete and perfect futures markets.

The actual consumer optimization problem in Eeckhout’s model does not involve state-dependent assets nor does it allow state-contingent transfers of income. If it were to allow this, as in a standard model of complete futures or insurance markets, then agents would never move. They would simply buy assets at the start of time that would pay them under a bad state in their city at a particular time, and such that they would pay under a good realization in their city. In other words, they would insure against the state of nature in their city. It is important to recognize that in this model there are two factors determining a worker/consumer’s productivity, namely the city-specific shock, and the externality in production induced by total population in the city.

The basic model of Duranton [Dur07] has consumers maximizing an intertemporal utility function subject to an intertemporal budget constraint, without facing uncertainty. However, once the detailed urban features are added (in Section V and Duranton [Dur06]), the model looks similar to Eeckhout’s at least in terms of the urban features. One simply needs some dependence of local prices (land rents or wages) on the state of nature. Then utility equalization implies that people will move depending on the state realization, but this movement disappears if one allows insurance.

There isn’t enough detail about the urban market in Duranton [Dur06, Dur07] to make specific
statements about how insurance would work, but the consumers in a city face uncertainty about employment due to the uncertainty about innovations in various industries, so similar insurance arguments should work if the details of the model are filled in.

Regarding contemporary developments in this literature, Behrens, Duranton, and Robert-Nicoud [BDRN+10] is a very interesting contribution that does not employ Gibrat’s law to obtain Zipf’s law. Using a static model, a number of stylized facts are matched. There is an asymmetric information/adverse selection component as well as a potentially insurable luck component in the model.

In general, we are inquiring whether moving or buying insurance is cheaper for the consumers in these models. Typically in these models if moving costs are positive, it makes sense for consumers to stay put and insure.

1.3.3 Criticism of the Literature

1.3.3.1 How Insurance Reduces Population Movement

So how might this insurance occur in practice? Let’s assume either that consumers cannot perfectly observe the technology shocks to cities, or moving has a small cost, or both.

- Self insurance. Since consumers can transfer consumption across time, and they know that shocks are i.i.d., then they can borrow or use their savings in bad times and save (or pay off their loans) in good, staying in the same city. In the literature, the intertemporal uncertainty faced by consumers does not show up in their objective function, whereas the possibility of self-insurance does not show up in the budget constraint. The earlier quote from Eeckhout seems to imply that this is allowed, but the formal statement of the consumer budget constraint makes it clear that this is not allowed. This type of insurance exploits the fact that for any given city, the shocks are i.i.d. over time. Empirically, the place to look for self-insurance is in the savings response to local employment shocks.
• Insurance markets. In all of these models, at each time the state of nature (the random shock to each production function for each city) is known to all and verifiable\textsuperscript{5} before consumers make their decisions about consumption bundles and location. So this is a perfect setting for a viable insurance market. An insurance firm can step in or the continuum of consumers can simply pool resources in each period, smoothing their consumption without changing location so it is independent of the state in their city. This type of insurance exploits the fact that at any given time, the shocks are i.i.d. across cities. Empirically, one place to look for insurance is a cross-country comparison of how varying benefits of unemployment insurance affect mobility in response to local employment shocks.

• Futures markets. Consumers formulate plans to sell labor and buy consumption commodity and housing contingent on every possible state in every time period. There is no empirical complement. We mention this for completeness.

Given that for Gibrat’s law to hold, the shocks to each city in each period must be “small” (see Eeckhout ([\textit{Eeco}4], p. 1447), it seems reasonable to think that insurance would yield higher consumer utility than movement, if moving costs are at all significant or if consumers cannot observe shocks to firms perfectly, and thus face even a small amount of uncertainty in their optimization problems.

For models in the literature, consumers will choose to insure instead of move when insurance is available. A common feature of both the models in the literature and the model we have presented is the prediction that people will move and not insure. A major difference between our model and the balance of the literature is clear: An advantage of our model is that it can explain endogenously the lack of insurance, whereas the other models in the literature implicitly assume that such markets, namely insurance or self-insurance (saving and borrowing), do not exist. The empirical investigation of the use of insurance as a substitute for migration, especially when consumer heterogeneity is taken into account, seems quite interesting as a topic for future

\textsuperscript{5}Thus, such models differ from models of human capital, for example, where verification is not a realistic assumption and thus insurance against fluctuations is not to be expected.
research.

But as a preview, we present preliminary work. We compare US data with analogous data for Belgium and Germany. For Germany and Belgium, we use data on municipalities, whereas for the US we use data on MSA’s. Please note that all of this data is therefore truncated. For Europe, we use the data from Soo [Soo05], who obtains it from http://www.citypopulation.de/. We provide summary statistics for all three data sets in table 3.

<table>
<thead>
<tr>
<th>Country</th>
<th>Year</th>
<th>Germany</th>
<th>Belgium</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>152684.8</td>
<td>62959.74</td>
<td>156903.7</td>
<td></td>
</tr>
<tr>
<td>Standard Error</td>
<td>19194.33</td>
<td>7372.42</td>
<td>15141.98</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>77486</td>
<td>39261</td>
<td>80537</td>
<td></td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>286632.5</td>
<td>61239.92</td>
<td>391062</td>
<td></td>
</tr>
<tr>
<td>Sample Variance</td>
<td>8.22E+10</td>
<td>3.75E+09</td>
<td>1.53E+11</td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td>20425</td>
<td>24791</td>
<td>50052</td>
<td></td>
</tr>
<tr>
<td>Maximum</td>
<td>3425759</td>
<td>446525</td>
<td>8008278</td>
<td></td>
</tr>
<tr>
<td>Count</td>
<td>223</td>
<td>69</td>
<td>667</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. City Population Summary Statistics for Germany, Belgium, and the US

We present in table 3 the estimates of the models for Germany and Belgium, to be used in conjunction with table 2 for the MSA estimates for the US. We note that for MLE, there is no general analog of t-statistics for each parameter estimate, just an overall measure of goodness of fit such as the KS statistic. The sample sizes are very different in table 2, resulting in very different log likelihoods as well as AIC and BIC statistics.

Likely insurance mechanisms are more developed and moving costs are higher for Europe compared with the US. Thus, one would expect deviations from Zipf’s law and lognormal models, but not from GEV, for Europe as compared to the US. We examine AIC and BIC ratios of lognormal to GEV for each country in table 4. What we find is that in 3 of the 4 cases, the GEV fit is better for Europe as opposed to the US.

For a more complete analysis, it would be desirable to regress the ratios in table 4 on proxies for moving cost and insurance mechanisms in each country for a larger sample of countries. That would give us a reading on which model performs better, but is beyond the scope of this paper.
1.3.3.2 **Possible Objections to the Criticism**

*We emphasize that the criticism we make is a purely theoretical point concerning models in the literature.* Whether or not agents in the real world actually insure or self insure against city-wide risk is not relevant to the question at hand. Our point is that in the theoretical worlds of these models, insurance or self-insurance of the sort discussed in the previous subsection is implicitly excluded. The reasons are not given or, more importantly, included in the model. If these factors, such as asymmetric information, are included in the model to explain insurance market breakdown, other competing forces driving agglomeration can be important; see for example Berliant and Kung (2010), where it is shown that adverse selection alone can generate agglomeration. In other words, this criticism of the internal structure of the models, for example when there is a non-zero moving cost, is that the consumers are not behaving rationally if they don’t insure or self insure.

Next we present a discussion of why insurance market breakdown is not natural in the context of the models. Again, this is not meant to be a statement about the real world, but rather about whether the exclusion of an insurance option for consumers in the models makes sense.

The usual cause of a breakdown of insurance markets is adverse selection, represented for example by cream-skimming on the part of insurance companies. In the models discussed here, the state is assumed realized and observable to all before decisions are made in a given time period. So there is no issue of adverse selection. But one can easily imagine variations of these models that incorporate some form of information asymmetry. It would not be natural for, say, only consumers to know the shock to the local economy, since the technology shock really affects firms. If only firms knew the realization of the shock before making their decisions, then consumers could draw inferences from firm behavior, or the consumers could self insure.
or insure. It is not clear what hidden information or hidden action on the part of consumers would cause an insurance market breakdown in this context, given that the shock is to firms’ technologies. It is natural to assume that amenities are observable.

One can imagine moral hazard at the city level with insurance markets, in that a city might try to claim a productivity level lower than the actual one so the residents can collect more insurance money. However, there are no local governments in the models in the literature to coordinate this, and the assumption is that local productivity is observable to all, including non-residents of the city, when they make their location decisions.

Another objection that could be raised is the commitment required on the part of consumers. In fact, commitment to a plan or contract is a requirement of models that feature self-insurance, insurance or futures markets generally. For example, a consumer might experience regret over the purchase of a long-term health insurance contract after the state of the world that tells them that they are healthy is realized. Or the insurance company might experience regret if the consumer turns out to be unhealthy. But they are committed to their contracts. In the models of the size distribution of cities, for example, one could begin the random process of technological change and at any point in time, allow insurance and commitment to begin. Then the population distribution will not change from that point on.

Self insurance through borrowing and saving requires a long term commitment to a plan. Insurance cooperatives or firms only require a one period commitment to stay in a city and work. The latter commitment problem can be solved with the following time line, a standard time line for insurance in the real world. First, people are in a city from last period. They make an insurance premium payment to the insurance company equal to the maximum possible income for a shock this period less the income workers received from work last period in the city. This will be “small” since the random shock is small, as explained in detail in appendix A.1. Then they work and the shock for this period is realized (timing here is not important). Then any insurance payment is made from the pool to obtain the average income. After that, the next period begins.
This way, people cannot receive income and then move without sacrificing their insurance. Since in all equilibria the utility levels in every city in Eeckhout’s model are the same, they must lose utility by moving and giving up insurance (the loss is their premium). Of course, one could then say that the insurance company could abscond with the money. But this stretches credulity.

One might easily object to even small moving costs or even a small amount of noise in consumer observations of shocks. Then what we present is another equilibrium, that yields exactly the same period by period utility as the equilibrium studied in this literature. This alternative equilibrium retains the initial distribution of consumers, and does not generate Zipf’s law.

Finally, there are costs associated with insurance contracts that, from the point of view of consumers, must be balanced against the cost of moving. Such costs involve lawyers and potentially complex transactions. Moreover, unemployment insurance might fulfill the role of explicit contracts. Self insurance does not suffer from these problems. But credit constraints could limit self-insurance. In any case, insurance does not need to be perfect. If there is substitution between insurance and mobility, the type of mobility needed to generate the various empirical distributions of city size can be upset.

But we emphasize again that although these various insurance market imperfections can cause insurance market breakdown, their inclusion in a formal model is necessary to ensure that consumers behave rationally when they don’t insure, and the consequences of their inclusion are far from obvious.

An alternative to insurance markets is self-insurance. Even if insurance markets are excluded from a model by assumption, for example because they are not observed in the real world, self-insurance must also be excluded and this exclusion justified.

In appendix A.1, we modify a model from the literature, Eeckhout [Eec04], to include insurance (as well as moving cost) and to prove our claims formally. This represents an example. We conjecture that the other models in the literature can be modified in a similar fashion.
1.4 Conclusions

We are making several related points:

- First, when a model, markedly different from those found previously in the literature, is constructed to explain a specific empirical phenomenon, the microeconomic, structural assumptions about individual behavior and markets must make sense. Here, there is a rather obvious problem that self-insurance and insurance markets are assumed not to be functional. Models in the literature feature city-level risk, and it is generally possible to insure against such risk through many vehicles, barring asymmetric information. The latter does not arise naturally in these models, since consumers are assumed to know the state of nature before making their location and consumption decisions.

- With time in the model, it is even possible to insure against aggregate risk through borrowing and saving.

- However, it is much more difficult to insure against extreme aggregate risk, so we propose such a model. Our model begins with microfoundations and delivers a different functional form for the size distribution of cities than has been used in the literature.

In summary, we first propose a model based on primitive assumptions, not designed to match any particular stylized fact (like the rank size rule), but rather capturing the following theoretical notion: Insurance is allowed, but consumers will never use it, as it is very costly. Instead, they move. The new model is based on extreme value theory and yields a functional form for the size distribution of cities different from the other models, and this prediction is empirically competitive with the ones in the literature. Then we advance a criticism of the literature based on the fact that a primitive assumption in previous work, that consumers cannot insure (either by borrowing and saving or by pooling resources) against the random productivity variable for each city that is observable to all. If insurance is allowed, there is another equilibrium of the model retaining the initial distribution of consumers where there is never any migration. Instead,
consumers insure against the risk, and the utility stream they obtain in this manner is the same as that in the equilibrium used in the literature. If there is any moving cost or residual uncertainty, the equilibrium used in the literature disappears.

Is insurance or self-insurance an important issue for the analysis of the size distribution of cities and city growth? The presence of insurance has no effect on our model, since it will never be taken up, and is simply prohibited in the other models in the literature. Thus, direct evidence regarding insurance or self-insurance is insufficient to distinguish between the models empirically. From the theoretical viewpoint, it makes no difference whether or not insurance is prohibited in our model, as the equilibrium is unchanged. But it makes a huge difference whether insurance is prohibited in other models, as the equilibrium with insurance and the equilibrium without insurance are vastly different. Other models from the literature that are modified to include insurance will not generate Zipf’s law or Gibrat’s law. It is in this sense that abstraction from consideration of insurance or self-insurance by other models in the literature is a first-order issue.

Future work includes testing further predictions of the model, for example the wage and rent distributions when transport costs for consumption commodities are introduced, and applying the model in new (but appropriate) contexts, such as finance (see Gabaix et al. (2003) for an application of Gibrat’s law to finance) or crop abundance (see Halloy (1999) for an application of the lognormal distribution to crop abundance).

Application to the size distribution of firms is of interest; see, for example, Axtell (2001) in the context of Zipf’s law or Gabaix (2011) more generally. Frequent churning might be expected more in firms than in cities. There are two issues with this idea. First, in an aspatial model, moving between firms is easy for workers, so our insurance critique will not apply to models using the lognormal or Pareto distributions, which therefore might be more appropriate. Second, we are using a competitive model since there is a continuum of firms in each city producing the same commodity and subject to the same productivity shock. The competitive assumption might not make as much sense in an aspatial model where productivity shocks are firm-specific, so
only one firm has the state of the art production technology.

Finally, an interesting direction for future research is to merge our model with that of Eaton and Kortum (2002). Such a model would be very complicated. As an alternative to this approach, adding an iceberg transportation cost to our model, as we have done in simulations, seems more worthwhile.
Chapter 2

A Scale-Free Transportation Network Explains City-Size Distribution

2.1 Introduction

Cities develop in relation to other cities rather than in a vacuum. What we consume in a city differs from what we produce in a city. The gap between the range and scale of production and consumption at city level is bridged by the transportation network, over which cities trade their products with others. The transportation network, in turn, does not coordinate cities uniformly. Some cities have only limited connections while others receive many links from cities across the country, both large and small, near and far away. The fate of city’s economy, and by extension its population size, is more or less conditioned by how it is positioned (inadvertently or otherwise) in the overall interurban network of cities and how accessible it is from others. We will show that the city-size distribution is the result of a particular class of network that our economy installs on itself for interurban trading purposes, namely, a scale-free network.

The existing literature’s treatment of the transportation network has been rather naïve and simplistic. Most existing models of city-size distribution implicitly or explicitly assume a completely isolated graph (figure 4(a)) or complete graph (figure 5(a)). Each node represents a city and a link represents a route available for shipment. The number inside a node counts its degree, i.e., the number of edges or routes each node has. Commodities cannot be shipped at all on a completely isolated graph, but they can be shipped anywhere in a single step from any city on a complete graph. Either way, the resulting equilibrium will be an even split of population among the cities, which does not match the actual city-size distribution. To explain the city-size distribution, we have sought a source of variation other than what the nexus of interurban rela-
(a) The United States according to completely isolated graph with the 50 largest cities.

(b) Completely isolated graph with 50 nodes.

Figure 4. Completely isolated graph

(a) The United States according to complete graph with the 50 largest cities.

(b) Complete graph with 50 nodes.

Figure 5. Complete graph

tionships has to offer. Some use a completely isolated graph (e.g., Eeckhout [Eco4]). Others such as Duranton [Duro6], Rossi-Hansburg and Wright [RHW07] or the New Economic Geography [FKV99] engage a complete graph as the transport structure, when in fact, transaction and/or communication between hub cities is much easier than between cities on peripheries. Behrens
et al [BMMS13] introduce a more lifelike representation of transportation cost in that the delivered price depends on a particular city pair. The price differential reflects monopolistic pricing rather than the underlying transportation network structure, which is still an (ex-ante) complete graph. The literature usually introduces a tiebreaker in the form of externalities, random growth, economies of scale or scope to replicate the actual city-size distribution.

In practice, transportation cost differs greatly depending on where you are and where you are headed. We will drop the assumption that our economy operates on a complete or completely isolated graph and see how much explanatory power network structure exerts as the engine of local economies of various sizes.

The transaction pattern between any two cities affects both the way cities are populated and the overall city-size distribution. Cities are tied together in various ways both topologically and economically. Some cities function as an intersection of major transportation routes and they trade and process commodities frequently in large volume. Others are less active in the interurban exchange of commodities. Differences among cities in terms of exchange patterns reverberate in the city-size distribution. Cities heavily interrelated to many others are likely to grow due to increased economic activities, whereas cities with sparse connections to a limited number of cities are liable to remain small in size. Those small cities, however, will not be completely wiped off the map.

2.1.1 CITIES ON A NETWORK

Intercity exchange patterns like figures 4(a) and 5(a) are best described by a network with cities as a set of vertices and traffic by edges as in figures 4(b) and 5(b). In this regard, network theory is indispensable when constructing a model of cities in the nationwide economy.

The recent seminal work by Barabási and Albert [BA99] has revitalized network theory. Classical network theory pioneered by Erdős and Rényi [ER59]’s model (ER network) cannot explain the emergence of a cluster or hub in a network, which we observe in most real social networks. In a classic random graph, each node is linked with an equal probability to any other and lacks dis-
tinctiveness, for the number of pre-existing links does not matter in forming a network. Barabási and Albert (BA) add a dynamic feature and preferential attachment to the classical random graph model so that the nodes are no longer ex-ante identical. Some nodes gather lots of links while others are wired to just a few. The model has been applied to many fields, including the emergence of web science, and has produced an improved description of the organization and development of networks. Most real-world networks have one thing in common: the resulting distributions of links are scale-invariant, that is, the distributions have fat tails. We can find nodes with an extremely large number of links rather easily with these networks compared to a classical random graph.

The city-size distribution shares the same pattern of scale invariance: the distribution of the 100 largest cities follows the same distribution as the one for the 1000 largest cities and so on, a property known as a power law, and in particular, Zipf’s law in the city-size literature. We expect that the degree of a city is positively related to its population. And for that reason, we imagine that our economy is based on a BA network rather than an ER network. This turns out to be correct, but selection of the appropriate network structure depends on exactly how node degree is related to city size. We will decode their relationship in section 2.3.8.

The urban economic application of network theory is in its very early stage of development and there is much room for advancement. Interaction between individual cities has not caught much attention so far. Our goal in this paper is to bring to the fore the interaction between transportation network structure and the city-size distribution. With this goal in mind we introduce (asymptotic) techniques from network theory and merge them with a tractable economic model in a new way. We do not intend this work to be the last word on this topic, but merely a suggestion of a first step into a bigger research program.

### 2.1.2 Some Transportation Networks Are Scale Free

Our economy operates on various modes of transportation and each mode comes with distinct network structures. Take a highway and airline network for example. Figures 6(a) and 7(a) are
schematic representations of the Interstate System and a typical airline route map for the 50 largest US cities. Apparently, a network composed of the Interstates does not share its structure with that of airlines. The Interstate will remain relatively intact when we take away New York, Houston and Cleveland. On the other hand, it would prove devastating if we did the same to the airline network (cf. [BB03]). More broadly, there is not much variance in the degree of nodes in the Interstate network, whereas the airline network has a limited number of heavily wired cities.

Figure 6.

Figure 7.
The BA network (figure 7(b)) explains the latter network better, as it follows a power law.

It should be noted, however, that what is geographically visible may not represent the real network that our economy relies on in effect. The Interstate network exhibits an ER-type topology as in figure 6. Nonetheless, the economy may operate a transportation network of a scale-free class on it. Shipment from Memphis has to go through St. Louis even if its final destination is Chicago. In this case Memphis is connected to Chicago in a single step rather than in two steps via St. Louis. For a carrier making Chicago-bound shipment from Memphis, St. Louis (a seeming layover node) is no different from the cornfield they pass through along the way (just a part of the edge), in that neither one of them add anything to the shipment. An economically relevant network is buried beneath the easily noticeable surface network and we do not want to confuse one with the other.

It is very important to note here a difference between the literature on dynamic social network formation and transportation networks. In the standard economics literature on social networks, for example Mele [Mel10] or Christakis et al [CFIK10], it is the individual agents, represented by nodes, who make decisions about forming links among themselves. In contrast, the nodes of a transport network are cities. Typically, it is not the cities or their agents who make decisions about forming links. Rather, it is another agent who controls an entire networks, for example the federal government in the case of highways or airlines in the case of an airline system.

2.1.3 The City-Size Distribution Is Scale Free Too

The city-size distribution has a distinct feature. Figure 8 plots the frequency of the city-size distribution from US Census 2000. It is only when we take the log of population (figure 8(b)) that the distribution exhibits resemblance to a familiar Gaussian distribution. Black and Henderson [BH03] and Soo [Soo05] explain how widespread scale-free distributions are in urban economics. Under the scale-free distribution, the arithmetic mean (Hillsboro, TX in figure 8)  

---

1 Scale-free distributions are commonplace in the socioeconomic realm. It seems that something of an additive nature prevails over natural phenomena, leading to a Gaussian distribution, and something of multiplicative nature (cf. [LSAo1]) is at work among socioeconomic phenomena, leading to a scale-free domain. We study the latter.
becomes less interpretive and the geometric mean (Sutton, NE) takes over the role of the average in the conventional sense.

Figure 8. Frequency plot of the city-size distribution. Dots are size proportionate. See table 7 for explanation of the cities selected in the figure. Data source: US Census 2000.
The fat-tailed distribution also makes its appearance on a map. Figure 9 illustrates the population density of each metropolitan and micropolitan statistical area (MSA and μSA, collectively referred to as Core Based Statistical Area, CBSA) in the United States in 2000. Most of the cities have a low density and are painted in blue; there are only few cities that are green and only two cities are colored in red. If the city-size distribution followed a Gaussian distribution or Poisson distribution with a large mean\(^3\), most of the cities should be green and only a few should be in blue or red. Just as for the airline network in figure 7(a), if we take away the ten largest US cities, we will leave more than a quarter of urban population unaccounted for.

Our main findings are as follows. City sizes are positively related to their degree. A city with a high degree has good accessibility to other cities. Reduced transportation cost makes the city’s product inexpensive and stimulates a large demand. As a consequence, the city creates large-scale employment. However, a marginal increase in degree contributes less to the city size as the degree increases. If a city is well-connected, then adding a new link to the city will not increase accessibility much because the city is already readily accessible from other cities through

\(^3\)As in the degree distribution of an ER network.
the existing grid.

We test implications of our model with Belgian and US data. The BA network leads to a result comparable to existing models, whereas the ER network fails to replicate the empirical city-size distribution. This confirms that the BA transport network is more consistent with reality.

The rest of the paper is organized as follows. In section 2.2, we will go over the two types of network structures mentioned above as a preamble to the next section, where we introduce and develop a model of spatial equilibrium with a transportation network woven into it. Particularly, in section 2.3.8, we will connect the network structure to the city-size distribution. In section 2.4, we verify the prediction of our model with data before we draw conclusions from our project in section 3.5.

2.2 Preliminaries

We will briefly review how ER and BA networks are built and examine the qualitative differences in terms of their degree distributions before we apply them to transportation networks.

2.2.1 ER Networks

The ER network is the simplest random graph of all. A pair of nodes are connected with a fixed connection probability. A completely isolated graph illustrated in figure 4 and complete graph illustrated in figure 5 are the special cases of the ER network where connection probability is zero and one, respectively.

The degree distribution of an ER network follows a Poisson distribution. The important feature is that the degree distribution is concentrated around its arithmetic mean\(^3\) and we rarely observe a city with an exceedingly large degree. All pairs of nodes share the same ex-ante connection probability, which leads to a small variance, and the network is egalitarian in that sense.

\(^3\)Recall that arithmetic mean does not mean much for scale-free distributions like the city-size distribution or a BA degree distribution.
2.2.2 BA Networks

The degree distribution of most real network structures does not follow a Poisson distribution. Rather, it follows a power law. This class of networks is called scale free. There are a number of proposed generative models that lead to power-law degree distributions (see Section VII of Albert and Barabási [AB02] for a review). To get a sense of how power-law type behavior emerges, consider the BA model [BA99] for example. Two major characteristics of BA model are growth and preferential attachment. The model sets off with a complete graph of a fixed number of nodes as a starting grid. New nodes with edges will be added sequentially to the existing network (growth).

As we can see from this mechanism, in general, older nodes are likely to gain an excessively large number of edges. The rich get richer because they are already rich (known as the Matthew effect). The rest of the nodes are merely mediocre in terms of degree. They are poor because they are already poor. This type of variance in degree hardly arises with an ER network. That is, New York City will not happen if the links are formed uniformly at random. Compare BA network figure 7(b) to ER network figure 6(b). BA network is not egalitarian, as connection probability depends on the number of acquired edges, which is path dependent. We shall also employ the network structure of Jackson and Rogers [JR07] that contains both the ER and BA types of networks as special cases. Details shall be provided in section 2.3.8.

2.3 Model

We propose a model where the trading costs of commodities among cities are explicitly specified. The city-size distribution is derived as a result of gains from trade and the underlying transport network configuration.
2.3.1 Location-Specific Commodities

There are \( J \) cities in the economy, with index \( j \). A city is defined as a geographic entity within which it produces the same commodity and from within which the geodesic paths (the shortest path on the network) to any other city in the country have the same length. If Adam and Beth both live in St. Louis, then they have the same shipping cost schedule to everywhere in the nation. We know they are in different cities if Adam pays a 10\% shipping charge to San Francisco and a 5\% charge to Minneapolis, whereas Beth pays a 10\% charge to San Francisco but an 8\% charge to Minneapolis. The endogenous population of city \( j \) is given by \( s_j \) and in total, there are

\[
\sum_{j=1}^{J} s_j = S
\]  

households in the economy. Each household supplies a unit of labor inelastically. City \( j \) produces consumption commodity \( c^j \) in a competitive environment. We assume that technology exhibits constant returns to scale and that one unit of labor produces one unit of commodity. In what follows a superscript denotes a city of production or origin, whereas a subscript denotes a city of consumption or destination.

The delivered price of commodity \( j \) in city \( i \) is denoted by \( p^i_j \). The value of marginal product \( p^i_j \cdot 1 \) coincides with the local wage \( \bar{w}^j \) in equilibrium:\footnote{Note that \( p^i_j \) denotes the mill price.}

\[
p^i_j = \bar{w}^j
\]

Consumer preferences are represented by a Cobb-Douglas utility function of the form \( u(c_i) = \frac{1}{J} \sum_{j=1}^{J} \log(c^j_i) \). The set of consumption bundles is constrained by the budget \( \bar{w}^i \geq \sum_{j=1}^{J} p^i_j c^j_i \).
2.3.2 Network Infrastructure and Delivered Price

The economy has a network infrastructure $\Gamma = (V,E)$, where $V = \{1, \ldots, J\}$ denotes the set of vertices representing each city and $E$ denotes a set of edges. For example a completely isolated graph in figure 4 is given by $\Gamma = ([1, \ldots, 50], \emptyset)$ and a complete graph in figure 5 by $\Gamma = ([1, \ldots, 50], \{|i,j| : 1 \leq i < j \leq 50\})$. All the traffic flow will follow $\Gamma$. We assume that the network is unipartite (i.e., there is a path between any pair of nodes) to avoid multiple equilibria. Whereas consumers in city $j$ can consume any commodity in the economy, they have to incur an extra iceberg transport cost to consume commodities brought in from other cities. Transportation cost piles up as a commodity travels from city to city along the path. To describe the exact transport cost structure, we define a metric $l_{ij} : V \times V \rightarrow \mathbb{R}_+$ to measure a geodesic length between node $i$ and $j$ given $\Gamma$. The delivered price of commodity $j$ shipped to city $i$ is given by

$$p_i^j = \tau l_{ij} p_j^j$$

(10)

where $\tau (\geq 1)$ marks the iceberg transportation parameter. We use the iceberg transport technology, standard in urban economics, for tractability reasons. If you dispatch $\tau$ units of commodity to your neighboring city, one unit of it will be delivered and the rest melts en route. The delivered price snowballs as the package travels from one city to another and the initial mill price is inflated by $\tau l_{ij}$ by the time the package reaches its final destination $l_{ij}$ steps over. We assume that all the links share the same value of $\tau$. The large fraction of transportation cost is a location-invariant fixed cost. Having $\tau$ dependent on each link will not add much to our analysis but will make our equilibrium analytically insolvable.

For detailed discussion, see McCann [McC05].
2.3.3 **Equilibrium**

Simple calculations yield the Marshallian demand for commodity $c_i$:

$$
\varphi_j^i(p^1_i, \ldots, p^J_i, w^i) = w^i (\tau^i_j p^j_i)^{-1} J^{-1}.
$$

The aggregate demand for commodity $j$ is the sum of demand from all the cities in the country:

$$
C_j(p, w) := \sum_{i \in V} s_i \varphi^j_i().
$$

Recalling that each household supplies one unit of labor inelastically and one unit of labor produces one unit of output, the commodity market $j$ clears when

$$
s_j = C_j(p, w) = \left( p^j_i \right)^{-1} J^{-1} \sum_{i \in V} s_i w^i
$$

The indirect utility function is given by

$$
v(p^1_i, \ldots, p^J_i, w^i) = \frac{1}{J} \sum_{j=1}^J \log \varphi^j_i() = \log w^i - \log J - \frac{1}{J} \sum_{j \in V} \log p^j_i - a_i \log \tau,
$$

where

$$
a_i := \frac{1}{J} \sum_{j=1}^J \langle l^j_i \rangle = \langle l_i \rangle
$$

is a remoteness parameter, or an average geodesic length from city $i$, where $l_i : j \mapsto l^j_i$. In what follows $\langle x \rangle$ denotes the average value of $x$. The parameter measures how hard it is to reach city $j$ from other cities in the economy. The higher the value is, the more remote the city is because we have to go through many links to get there. We will explore the role of accessibility later.

Free mobility of consumers implies

$$
v(p^1_i, \ldots, p^J_i, w^i) = v(p^1_i, \ldots, p^J_i, w^j)
$$

---

6 This expression may seem incredulous at first, for it does not include $\tau$. A large $\tau$ discourages demand but it also means that firms have to ship more commodities. A large portion of shipment will melt on its way. They cancel each other in equilibrium. This propitious cancellation may not occur with other preference specifications.
for all $i, j \in V$ in equilibrium.

The equilibrium $(s_1, \ldots, s_J; p^1, \ldots, p^J; w^1, \ldots, w^J)$ satisfies (8), (9), (11) and (13). Utility equalization (13) leads to

$$\log p^i - \log p^j = (a_i - a_j) \log \tau. \quad (14)$$

Equation (14), together with (11), implies $s_j = \tau^{a_i-a_j} s_i$. With the population condition (8), we obtain the city-size distribution

$$s_i = \frac{S}{\tau^{a_i} \sum_{j \in V} \tau^{-a_j}}. \quad (15)$$

2.3.4 How Does a Network Break Symmetry?

An obvious implication of (15) is that cities with better accessibility have larger equilibrium population. Naturally, we are tempted to conclude that the entire population will collapse into the city with the best accessibility and the rest of the cities will be completely vacated. As it turns out, this is not the case. The city-size distribution will not become degenerate. Let us break down (15) both mathematically and economically to see why.

First, let us recast the relationship (15) to explore how accessibility translates to the population of a city. We can rewrite (15) as $s(a_i) = \langle s \rangle \tau^{-a_i} / \langle \tau^{-a} \rangle$, where $\langle s \rangle := S/J$ is a base city size and $\langle \tau^{-a} \rangle := \sum_j \tau^{-a_j} / J$ gives the average of $\tau^{-a_j}$. The city size spreads around the canonical size $\langle s \rangle$. A better accessibility (i.e., small remoteness value $a_i$) contributes to the city by augmenting the baseline size $\langle s \rangle$ by a factor of $\tau^{-a_i} / \langle \tau^{-a} \rangle$. The multiplier is large when $\tau^{-a_i}$ is greater than the national average $\langle \tau^{-a} \rangle$ and vice versa. Furthermore, the multiplier grows more than proportionally.
as the city’s accessibility improves as can be seen in figure 10. The multiplier $\tau^{-a_i}$ is monotone decreasing and convex in $a_i$. Does this mean New York City sweeps away all the population off the rest of the cities? — Not really. And it calls for an economic exposition of (15) to see why.

Although restricted accessibility of a city raises its delivered prices, demand for its produced commodity does not cease to exist. Eliminating a commodity from the basket will punish consumers a lot. They appreciate variety and missing a single variety will push the utility level down to negative infinity. Workers in a poorly connected city will have to pay a high price for imported commodities due to a poor network infrastructure, but they are compensated with a high nominal wage, as indicated by the wage (9) and utility equalization (14). These two equations imply that the mill price (and ultimately, the nominal wage) is positively related to the average geodesic length $\langle l_i \rangle$ from city $i$ in equilibrium, i.e., a sparsely connected city has a high mill price. The prices adjust to make it worth living in small cities in equilibrium. The scale of local production is small, but each commodity is sold high to make up for an increased cost of living due to remoteness and the resulting costly transport.

Variance in city sizes is solely due to the structure of the network. The above-mentioned trade-off entails two counteracting forces. The agglomerative force is heterogenous accessibility, which tends to spread out the city-size distribution. The dispersion force is preference for variety, which tends to push the distribution back to a collection of equal-sized cities.

There are alternative ways to derive city size with a tractable economic model, particularly for the dispersion force. In this model, location-specific commodity production drives dispersion, as a bundle of all goods is desired by consumers. An alternative model would use another natural dispersive force, say housing or land markets. If we had just a few produced commodities (say one for illustration), then Starrett’s Spatial Impossibility Theorem (Fujita and Thisse [FT02], Ch.2) applies, and we would have an autarkic equilibrium where no commodity is transported. Yet another alternative is to introduce a congestion externality, but then the model begins to look more complicated and, at the same time, arbitrary.

---

7Starrett’s Theorem makes no assumption about the transport network or transport cost.
Obviously, this trade-off disappears and there will be no variance in city sizes if the agglomerative force is removed. This can happen when shipment becomes costless (to be discussed in proposition 2.3.1) or network structure becomes redundant, that is, if it turns into a complete graph. Although we introduced a location-specific technology, commodities are symmetric. Technology is linear everywhere. Consumer preferences are identical and they put the same weight on each commodity. If we take the network structure out of the equation, the resulting equilibrium is such that all the cities share the same size \( s \) and every household consumes an equal portion of all the commodities available.

2.3.5 **Transportation Cost Skews the City-Size Distribution**

Along with remoteness \( a_i \), transportation cost \( \tau \) plays a leading role in the determination of the city-size distribution. Depending on its magnitude, \( \tau \) can nullify or amplify the influence of a network structure over the economy. Figure 10 compares the relationship between accessibility and the city-size distribution under different transportation costs.

In the extreme situation where shipment is free (\( \tau = 1 \)), all the cities will be of an equal size regardless of the network structure. The city size \( s(a_i) \) becomes constant against \( a_i \) (see the blue line in figure 10). The network becomes a complete graph in effect, because the delivered price will be the same no matter how long the geodesic length is. For \( \tau > 1 \), city size (15) becomes a strictly convex function of remoteness.

The transportation network \( \Gamma \) starts to sink in as \( \tau \) grows. A large \( \tau \) implies that the geodesic length exerts a more dominant influence on the size of a city. With a small value of \( \tau \), a city with good accessibility does not distinguish itself
well from other cities because the effect of path length is limited due to low transportation cost. On the other hand, if shipping is costly, a city with a good accessibility benefits from a low \( a_i \) value because high transportation cost amplifies the effect of accessibility. In other words, a high transportation cost reveals the network structure and projects the network \( \Gamma \) onto the city-size distribution in a more pronounced, clear-cut manner than with a low transportation cost. As a result, holding the remoteness distribution constant, large \( \tau \) skews the city-size distribution and makes the emergence of disproportionately large hubs more likely. To measure how the cost of transportation \( \tau \) bends the city-size distribution, consider a measure

\[
D(\tau) = \frac{s(a_H) + s(a_L)}{2} - s\left(\frac{a_H + a_L}{2}\right),
\]

where \( a_H \) and \( a_L \) are the highest and lowest remoteness of a given network. The first term is the average of the smallest and the largest city whereas the second term is the city size of average remoteness. For a given distribution of remoteness \( a_i \), \( D(\tau) \) measures the convexity of \( s(a_i) \), which gauges how spread out the distribution of city size \( s(a_i) \) is for each \( \tau \). See figure 11. When \( \tau = 1 \), \( s(\cdot) \) lays flat and \( D(\tau) = 0 \). As \( \tau \) grows, \( s(\cdot) \) bends more and \( D(\tau) \) grows accordingly as can be seen in figure 10.

We confirm the observation above as follows:

**Proposition 2.3.1:** Transportation Cost Skews the City-Size Distribution

*Suppose that the economy has a unipartite network \( \Gamma \). The city-size distribution \( s_i \) is a convex function of remoteness \( a_i \) for \( \tau \geq 1 \). Moreover, the degree of convexity measured by the size difference \( D(\tau) \) between the city of average size and the city of average remoteness increases with \( \tau \).*

*Proof.* See appendix A.3. \( \square \)

### 2.3.6 Geodesic-Length Distribution

The city-size distribution (15) depends on the distribution of remoteness (12), which, in turn, rests on the distribution of geodesic length. While most of the research on network topology is
focused on mean intervertex distance ([NSW01], [FFH04], [ZLG+09]), what we need here is the geodesic length between individual nodes. Mean intervertex distance comes in handy when we gauge how efficient a network is, but we are not here to see if the transportation network that our economy relies on is optimally configured (that would be another paper). We would like to derive the city-size distribution, not the average size of cities or the remoteness thereof.

There is not much research that looks into the geodesic length between each pair of nodes. At the time of writing, the analytical form of geodesic length between individual nodes is yet to be discovered\(^8\). There is an attempt to track down the geodesic length by guessing the analytical form from sequentially generated, fractal-like networks reverse-engineered from a Pareto degree distribution ([DMO06]), which we cannot use because our distribution (21) is not a Pareto distribution.

Hołyst et al [HSF+05] take a different approach to derive an intuitive solution for a wide range of network types. They measure the expected geodesic length between any pair of nodes \(i\) and \(j\) as follows:

\[
l^i_j = A - B \log(k_ik_j),
\]

where \(A := 1 + \log(J(k))/\log \kappa\) and \(B := (\log \kappa)^{-1}\). The number \(k_i\) denotes the degree of node \(i\). Rearrange the nodes so that we have a tree with node \(i\) as its root. The average number of children is called an average branching factor and denoted by \(\kappa\). For more details see appendix A.4.

Although [HSF+05] does not provide a formal proof of (16), but rather is based on a heuristic, it appears to be the best we can do given the current state of network theory. Zhang et al [ZLG+09] provide an analytical background for the mean intervertex distance for a special case. We hope that its extension to individual distances will become available in the near future.

Meanwhile, (16) proves to be quite useful in translating a network structure into economic context without loss of generality. A path length is a global property whereas a degree is a local property. We cannot compute the individual geodesic path unless we compare all the possible

\(^8\)The one for the average intervertex separation has already been brought out into the open. Cf. [NW99], [NMW00], [ZLG+09].
paths between a city pair of interest and pick the shortest one, which calls for a systemic search all across the board. The geodesic path thus obtained is too specific to the particular network in question and does not have wide implications beyond the specific network itself. Degree is much easier to compute because we do not have to launch a nationwide search for it, and the degree distribution is readily available for a wide range of networks. Equation (16) succinctly writes a global property (a path length) in terms of the analytically manageable local property (a degree). It implies that the path length will be short if your city and/or your destination city have many edges to choose from to begin with and/or to end with. This abundance in selection should save you from being thrown to circuitous paths, and vice versa when your degree is small. Absent this conversion of the global property into the local property, we would not be able to describe a general relationship between degree and city size, when in fact, there is an obvious symbiotic interaction between them waiting to be investigated.

2.3.7 City-Size Distribution

From (16), remoteness (12) is written as

\[ a_i(= \langle l_i \rangle) = A - B \log k_i - B\langle \log k \rangle. \]  \hspace{1cm} (17)

We observe that accessibility improves as a city acquires more edges, but only on the logarithmic order. Taking the log of (15), we have

\[ \log s_i = \log S - (A - B \log k_i - B\langle \log k \rangle) \log \tau - \log \left( \sum_j \tau^{-a_j} \right). \]
The last term is approximated by \( \log J - \langle a \rangle \log \tau \) so that
\[
\log s_i = \log \langle s \rangle + B \log \tau \left( \log k_i - \langle \log k \rangle \right). \tag{18}
\]

A couple of observations are in order. The equation above answers two questions concerning the relationship between a network structure and a system of cities. The first one is “Does construction of an edge boost the local economy?” The answer is “Apparently.” The second, and more interesting question is “How so?” The answer is twofold.

In terms of a linear scale, (18) can be rewritten as
\[
s_i = \langle s \rangle \gamma B \log \tau \left( \frac{k_i}{\gamma} \right),
\]
where \( \gamma = \prod_{i=1}^{J} k_i^{1/J} \) is the geometric mean of the degree. It indicates that city size is anchored around the base city size \( \langle s \rangle \) multiplied by the deviation \( \frac{k_i}{\gamma} B \log \tau \). If a city has a large degree, then its size becomes larger than the standard city size by a factor of \( \frac{k_i}{\gamma} B \log \tau \) and vice versa for a city with a small degree. The city size coincides with the cornerstone size of \( \langle s \rangle \) exactly when its degree matches the national (geometric) average.\(^{10}\) The deviation is amplified as shipment becomes costly, which, in turn, confirms our observation made in proposition 2.3.1.

We also note that adding an edge to a city increases its size, but the change in size is inversely proportional to the current degree provided \( B \log \tau < 1 \). If city \( i \) is highly wired already, then the introduction of a new edge to city \( j \) does not add much to city \( i \). The geodesic length to city \( j \) is already short before the establishment of the new edge. You can go to many cities in a single step and city \( j \) is likely to be linked to at least one of those many neighboring cities already, making the geodesic length to city \( j \) just two. The added edge will only reduce the geodesic length by

---

\(^{9}\) Let \( \vec{a} := (a_1, a_2, \ldots, a_J) \) and \( \langle \vec{a} \rangle := (\langle a_1 \rangle, \langle a_2 \rangle, \ldots, \langle a_J \rangle) \). The Taylor series expansion about \( \vec{a} = \langle \vec{a} \rangle \) tends to
\[
\log \left( \sum_i \tau^{-a_i} \right) = \log \left( \sum_i \tau^{-\langle a \rangle} \right) + (\vec{a} - \langle \vec{a} \rangle) \cdot D \log \left( \sum_i \tau^{-a_i} \right) \bigg|_{\vec{a}=\langle \vec{a} \rangle} + O \left[ (\vec{a} - \langle \vec{a} \rangle) \cdot (\vec{a} - \langle \vec{a} \rangle) \right]
\]
\[
\rightarrow \log J - \langle a \rangle \log \tau,
\]
by the law of large numbers.

\(^{10}\) This examination begs one question: If my city has the average number of edges, is my city larger or smaller than the national average in size? The answer is “larger”. Since transportation cost and the branching factor are both greater than one, \( \frac{\log k_i}{\log \tau} \) is positive. Plus, the geometrical mean is smaller than the arithmetic mean. To score a national average \( \langle s \rangle \) you only need \( \gamma \) edges. It should be noted, however, that in a scale-free world, arithmetic mean does not carry much information. The lognormal is the new normal (or any heavy-tailed distribution is for that matter) and the geometric average is the new average in this world as we saw in figure 8(b).
one. On the other hand, if the current degree of city $i$ is low, then the link to city $j$ will not only reduce the geodesic length to city $j$ greatly but also reduce the geodesic lengths to the cities in city $j$’s neighborhood. Consequently, city $i$ will see significant reduction in its average geodesic length.

Based on the degree-size relationship (18), our main theoretical result gives the city-size distribution as follows:

**Proposition 2.3.2: City-Size Distribution**

Suppose that the economy has a unipartite network $\Gamma$ with the associated degree distribution $G(k)$. The city-size distribution of this economy follows the distribution function $F(s)$, defined by

$$F(s) = G(k(s)), \quad (19)$$

where $k(s) := \gamma(s/\langle s \rangle)^{\frac{\log \kappa}{\log \tau}}$. Its probability density function (PDF) is

$$f(s) = k'(s)g[k(s)] = \frac{\log \kappa}{\log \tau}k(s)s^{-1}g[k(s)], \quad (20)$$

where $g(\cdot)$ denotes the PDF of degree $k$.

Since the transport cost and average branching factor only come into the equation in the form of a quotient of their logarithmic values, $\frac{\log \kappa}{\log \tau}$, we will denote this by $\delta$ for estimation purposes, in which case, (20) becomes $f(s) = \gamma\delta(s)^{-\delta}s^{\delta-1}g[k(s)]$. As we have already seen a small $\delta$ stretches out the distribution and a large $\delta$ does the opposite.

**2.3.8 City-Size Distribution under Different Network Systems**

Now that we have the city-size distribution based on the city’s degree, we can make our predictions based on different transport network structures. There are two network models of particular interest: ER and BA networks.

Note that empirical determination of the transport network relevant to the formation of a
system of cities is a tough job. The task at hand is to find a network that is consistent with the real city-size distribution (and we have already discarded complete and completely isolated networks in section 2.3.4). The most consistent network structure will give us a clue as to the shape of a network that is germane to the formation of cities.

Jackson and Rogers [JR07] constructed a degree distribution of a directed\textsuperscript{11} dynamic network as follows:

\[ G(k) = 1 - \left( \frac{k_0 + rm}{k + rm} \right)^{1+r} \quad \text{for} \quad k \geq k_0, \quad (21) \]

where \( k_0 \) denotes an in-degree with which an entering node is endowed. This value is shared across all the nodes. The ratio of the number of links formed by an ER-like random connection and a BA-like network-based connection is given by \( r \), and \( m \) is the average out-degree of a node. Five PDF’s of (21) are depicted in figure 12 as a visual cue. In the figure parameter \( r \) ranges from .01 (over 99\% network-based and less than 1\% random links) to 100 (the other way around). A predominantly random PDF (with large \( r \)) tapers off quickly whereas a mostly network-based PDF (with small \( r \)) only gradually dissipates with degree. We expect that our economy operates with a small \( r \). In what follows we refer to in-degree as the degree unless otherwise stated. BA network’s degree distribution is (21) with \( r = 0 \), in which case, (21) turns into a Pareto distribution. ER network calls for \( r \rightarrow \infty \), in which case (21) is no longer well defined and the degree distribution turns into an exponential distribution.\textsuperscript{12}

\textsuperscript{11}Commodities can flow either way on an edge. We take an arrowhead on a directed edge just as a decorative memorabilia indicating from which end the edge was constructed, but nothing more. We represent degree distribution by an in-degree distribution. It is impossible to tell different networks apart with an out-degree distribution due to the way a network is constructed in [JR07]. Any network comes with a degenerate out-degree distribution.

\textsuperscript{12}The original ER network [ER59] comes with a Poisson degree distribution rather than an exponential degree distribution. The differences in the distribution arise from the way the network is constructed: [JR07] is dynamic, whereas [ER59] is static.
What is left to do is write the mean branching factor \( \kappa \) in terms of other parameters in (21) before we can fully identify the city-size distribution.\(^{13}\) The actual mean branching factor cannot be computed until after the network is formed. Holyst et al [HSF+05] provide a good approximate to \( \kappa \):

\[
\kappa = \frac{\sum_{k=1}^{J} k g(k)}{\sum_{x=1}^{J} x g(x)} - 1 = \frac{\sum_{k} (2k - 1) G(k)}{\sum_{x} G(x)} - 1 = \frac{\mu_k^2 + \sigma_k^2}{\mu_k} - 1,
\]

where \( \mu_k \) and \( \sigma_k^2 \) denote the mean and variance of \( k \), respectively. For details, see appendix A.5.

While [JR07] is microfounded and sufficient to generate a fat-tailed degree distribution, it is not necessarily the only degree distribution which a BA network gives rise to. There is a chance that our economy’s transportation network may have come around from a different mechanism than [JR07]. In this regard we experimented with other fat-tail distributions as a candidate degree distribution along with (21). In particular, we tested lognormal and generalized extreme value (GEV) distributions for use as a degree distribution. To our knowledge, these degree distributions are not yet microfounded.

### 2.4 Empirical Implementation

Now that the model with an explicit transport system is at the ready, we will pitch it against the actual city-size distributions to identify what class of network governs the city-size distribution. By and large the results are in full support of our initial inkling that a scale-free network explains the city-size distribution but ER or other network structures commonly adopted do not.

All told, we have four sets of data on our plate: Belgium, Metropolitan Area (MA), CBSA and Places.\(^{14}\) Descriptive statistics for each data set are in table 7.

---

\(^{13}\)The branching factor is not a free parameter and it cannot be directly estimated from the data, because the estimation algorithm will either explode or create indeterminacy. It is dependent on the shape of the network, which, in turn, is characterized by the other parameters via (22).

\(^{14}\) The Belgian data is provided courtesy of Soo [Soo05] and the remainder are from US Census 2000. For definitions of MA and CBSA, see <http://www.census.gov/population/metro/about/> and for Places, see <http://www.census.gov/geo/reference/gtc/gtc_place.html>. We thank Jan Eeckhout for sharing his data used in [Eec04].
The Belgian data is included to see if our model’s predictive value is subject to both the area and population size of a country under study. (It was not.) MA and CBSA are a popular choice in the literature. The smallest unit of measurement is a county and they suffer from data truncation ([Eeco4]). Places have the finest unit of measurement and are free of truncation. We tested the following five distributions against them: ER/BA, BA, lognormal, GEV and the degenerate distribution. The first two distributions are estimated in three ways: maximum spacing estimation (MSE), minimum Kolomogorov-Smirnov estimation (minKS) and maximum likelihood estimation (MLE), and the remainder in MSE.

In what follows a hat on parameter $x$ indicates its estimate, $\hat{x}$. 
<table>
<thead>
<tr>
<th>Data</th>
<th>Belgium</th>
<th>MA</th>
<th>CBSA</th>
<th>Places</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data size $J$</td>
<td>69</td>
<td>276</td>
<td>922</td>
<td>25,358</td>
</tr>
<tr>
<td>Total urban population $S$</td>
<td>4,344,222</td>
<td>225,981,679</td>
<td>261,534,991</td>
<td>208,735,266</td>
</tr>
<tr>
<td>Population covered</td>
<td>42.38%</td>
<td>80.30%</td>
<td>92.93%</td>
<td>74.17%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Largest city</th>
<th>Antwerp</th>
<th>New York CMSA</th>
<th>New York MSA</th>
<th>New York city</th>
</tr>
</thead>
<tbody>
<tr>
<td>Largest size</td>
<td>446,525</td>
<td>21,199,865</td>
<td>18,323,002</td>
<td>8,008,278</td>
</tr>
<tr>
<td>City near arithmetic mean</td>
<td>Genk</td>
<td>Oklahoma, OK MSA</td>
<td>Green Bay, WI MSA</td>
<td>Hillsboro city, TX</td>
</tr>
<tr>
<td>Arithmetic mean</td>
<td>62,960</td>
<td>818,774</td>
<td>283,661</td>
<td>8,232</td>
</tr>
<tr>
<td>Median city</td>
<td>Beringen</td>
<td>Anchorage, AK MSA</td>
<td>Hinesville-Fort Stewart, GA MSA</td>
<td>Harristown village, IL</td>
</tr>
<tr>
<td>Median size</td>
<td>39,261</td>
<td>259,600</td>
<td>71,800</td>
<td>1,338</td>
</tr>
<tr>
<td>Smallest city</td>
<td>Arlon</td>
<td>Enid, OK MSA</td>
<td>Andrews, TX µSA</td>
<td>New Amsterdam, IN</td>
</tr>
<tr>
<td>Smallest size</td>
<td>24,791</td>
<td>57,813</td>
<td>13,004</td>
<td>1</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>61,240</td>
<td>1,968,621</td>
<td>974,190</td>
<td>68,390</td>
</tr>
<tr>
<td>Skewness</td>
<td>4.183</td>
<td>6.682</td>
<td>10.98</td>
<td>75.53</td>
</tr>
<tr>
<td>City near geometric mean</td>
<td>Mouscron</td>
<td>Huntsville, AL MSA</td>
<td>Sunbury, PA µSA</td>
<td>Sutton city, NE</td>
</tr>
<tr>
<td>Geometric mean</td>
<td>50,809</td>
<td>342,844</td>
<td>94,373</td>
<td>1,447</td>
</tr>
<tr>
<td>Mean of log(s)</td>
<td>10.84</td>
<td>12.75</td>
<td>11.46</td>
<td>7.278</td>
</tr>
<tr>
<td>Standard deviation of log(s)</td>
<td>.5697</td>
<td>1.119</td>
<td>1.191</td>
<td>1.754</td>
</tr>
<tr>
<td>Skewness of log(s)</td>
<td>1.498</td>
<td>1.048</td>
<td>1.187</td>
<td>.2091</td>
</tr>
</tbody>
</table>

Table 5. Descriptive Statistics. The statistics above the line (shaded in blue) are related to a linear scale and the below (shaded in green) are related to a log scale. Mean of log(s) is same as the log of geometric mean.
2.4.1 Estimation Methods Employed

The first choice is to go for MLE, which does not work with (21). The likelihood function is monotone increasing in \( k_0 \). As a workaround to MLE, we calculated the estimates by MSE. While its use is limited in the city-size literature so far especially when compared to MLE, it is more robust and easier to handle than MLE. The problem we have with MLE is exactly the one exemplified in Ranneby [Ran84] and we used his solution. The MSE estimator maximizes the geometric mean of the gap or step between two adjacent CDF’s

\[
F(s_i; \theta) - F(s_{i-1}; \theta),
\]

where \( \theta \) is a vector of parameters to be estimated and data sequence \( s \) is rearranged in the ascending order \( s_1 \leq s_2 \leq \cdots \leq s_J \).\(^{15}\) The idea here is to split the interval \([0, 1]\), the range of a CDF, in \( J \) steps in the way that none of the assigned \( F(s_i; \theta) \) will create a disruptively large gap with its neighbors and the gaps should be evenly spaced as much as possible on the logarithmic scale. Maximizing the arithmetic mean does not work here because it will always be \( 1/J \) no matter what estimates we toss in. This actually works as a ap on our geometric mean in turn, by Jensen’s inequality. Thus, we can safely rule out the possibility that the maximand tends to infinity, which is exactly the reason why we had to discard MLE. For more on MSE, see appendix A.6.

2.4.2 A Scale-Free Transportation Network Explains the City-Size Distribution

Estimation with four different data sets unanimously chooses BA over ER as the underlying transport network in our economy. We report our results in table 6 and figures 13 to 16 along with other distributions.

\(^{15}\) The first and last gap are defined by \( F(s_1; \theta) - F(-\infty; \theta) \) and \( F(\infty; \theta) - F(s_J; \theta) \) each.
| Data          | Distribution                        | ⟨log LH⟩ | KS     | ⟨log step⟩ | geo/arith | |θ| | BIC  | AIC | r |
|---------------|------------------------------------|----------|--------|------------|-----------|-----|-----|------|-----|----|
| Belgium       | Lognormal (Eeckhout)               | -11.69   | .1986  | -5.266     | .05166/.01449 | 2   | 1621| 1617 |
| Belgium       | GEV (Berliant & Watanabe)          | -11.40   | .1122  | -4.981     | .006870/.01449 | 5   | 1594| 1583 |
| Belgium       | Complete Graph (de facto)          | -∞       | .6812  | -∞         | 0/.01449    | 1   | ∞   | ∞    |
| Belgium       | ER/BA (Jackson & Rogers)           | -11.47   | .1348  | -5.072     | .006268/.01449 | 5   | 1604| 1593 | .002745 |
| Belgium       | ER (Jackson & Rogers)              | -11.49   | .1766  | -5.086     | .006185/.01449 | 4   | 1603| 1594 |
| MA            | Lognormal (Eeckhout)               | -14.28   | .1036  | -6.232     | .001996/.003623 | 2   | 7891| 7884 |
| MA            | GEV (Berliant & Watanabe)          | -14.13   | .04334 | -6.089     | .002267/.003623 | 5   | 7828| 7810 |
| MA            | Complete Graph (de facto)          | -∞       | .7935  | -∞         | 0/.003623    | 1   | ∞   | ∞    |
| MA            | ER/BA (Jackson & Rogers)           | -14.17   | .06102 | -6.134     | .002168/.003623 | 5   | 7852| 7834 | .001154 |
| MA            | ER (Jackson & Rogers)              | -14.21   | .1057  | -6.173     | .002084/.003623 | 4   | 7860| 7851 |
| CBSA          | Lognormal (Eeckhout)               | -13.05   | .09402 | -7.548     | .0005270/.001085 | 2   | 2.407e+04 | 2.4063e+04 |
| CBSA          | GEV (Berliant & Watanabe)          | -12.91   | .02606 | -7.409     | .0006056/.001085 | 5   | 2.384e+04 | 2.382e+04 |
| CBSA          | Complete Graph (de facto)          | -∞       | .8362  | -∞         | 0/.001085    | 1   | ∞   | ∞    |
| CBSA          | ER/BA (Jackson & Rogers)           | -12.95   | .05922 | -7.449     | .0005819/.001085 | 5   | 2.391e+04 | 2.389e+04 | .0004526 |
| CBSA          | ER (Jackson & Rogers)              | -13.29   | .1762  | -7.794     | .0004121/.001085 | 4   | 2.454e+04 | 2.452e+04 |
| Places        | Lognormal (Eeckhout)               | -9.258   | .01895 | -8.840     | 0/3.944e-05  | 2   | 4.696e+05 | 4.696e+05 |
| Places        | GEV (Berliant & Watanabe)          | -9.254   | .008847| -8.836     | 0/3.944e-05 | 5   | 4.694e+05 | 4.693e+05 |
| Places        | Complete Graph (de facto)          | -∞       | .8342  | -∞         | 0/3.944e-05 | 1   | ∞   | ∞    |
| Places        | ER/BA (Jackson & Rogers)           | -9.268   | .02198 | -8.849     | 0/3.944e-05 | 5   | 4.701e+05 | 4.700e+05 | .000317 |
| Places        | ER (Jackson & Rogers)              | -9.392   | .1134  | -8.974     | 0/3.944e-05 | 4   | 4.764e+05 | 4.763e+05 |
| Places        | Lognormal as Degree Dist.          | -9.258   | .01896 | -8.840     | 0/3.944e-05 | 4   | 4.696e+05 | 4.696e+05 |
| Places        | GEV as Degree Dist.                | -9.255   | .01159 | -8.836     | 0/3.944e-05 | 5   | 4.694e+05 | 4.694e+05 |

**Table 6. Model Comparison**

Row color corresponds to the line colors in figures 13 to 16. ▲ denotes a statistic the higher value of which indicates a better fit and ▼, the other way around. ⟨log LH⟩ denotes the average of the log of likelihood values, KS denotes the Kolomogorov-Smirnov statistic, ⟨log step⟩ measures the geometric mean of the step $F(s;\theta) - F(s_{i-1};\theta)$ in logarithms. Geo/arith measures the ratio between geometric mean and arithmetic mean of the step. The closer the geometric mean is to the arithmetic mean, the better the fit is. It is zero for Places due to multiple cities having the same size. |θ| counts the number of parameters. BIC and AIC stand for Bayesian and Akaike Information Criteria for detecting overfitting. **Boldface with white foreground** marks the winner and **boldface with black foreground** denotes the runner-up among the five distributions tested.
Figure 13. Model Comparison (Belgium)

Figure 14. Model Comparison (MA)
Figure 15. Model Comparison (CBSA)

Figure 16. Model Comparison (Places)
Figure 17. MSE for Belgium (left) and MA (right)
Figure 18. MSE for MSA (left) and Places (right)
ER/BA in the table corresponds to (21). We left the estimated distribution functions for ER/BA in figures 17 and 18.

As low values of \( \hat{r} \) indicate, edges are formed predominantly through networking rather than by random selection. We cross-checked estimates with minKS and MLE\(^{16} \) and we obtained a similar result. To be doubly sure of our findings, we ran estimation with \( r \to \infty \). ER in table 6 lists the statistics with \( r \to \infty \). The statistics of ER seem to be comparable with other distributions except that the estimated transportation cost is unreasonably high. A one-dollar pen will cost more than the US GDP five towns over on the purely random network.\(^{17} \) Thus, we conclude that a scale-free transportation network explains the city-size distribution but a scale-variant network does not.

Estimated \( \hat{\delta} \) ranges from .9911 to 2.536.\(^{18} \) As we discussed in reference to (18) we confirm that in most cases, the impact of adding an edge on city size wears off as degree itself becomes saturated (it cannot exceed \( J - 1 \)), or put differently, New York has more edges, size for size, than any other cities as it takes more edges to raise the city size as the city grows further.

We ran MSE with three other distributions representative of the existing city-size models to compare with our model. Eeckhout \( \text{[Eco04]} \)'s model leads to a lognormal distribution and Berliant and Watanabe \( \text{[BW14]} \) predict a GEV distribution as the city-size distribution. A complete graph will result in a degenerate probability distribution. The BA economy fits comfortably into the circle of existing testable models based on all the statistics we computed in table 6 (usually coming in second on all fronts except for Places).

In addition we put two other fat-tailed degree distributions to the test. The results (the last two rows in table 6) seem to indicate that the network formation does not necessarily have to be

---

\(^{16}\)With \( k_0 \) fixed at zero to prevent explosion. MSE and minKS point \( \hat{k}_0 \) towards zero.

\(^{17}\)There is not enough variance in the ER degree distribution, certainly not power-law type behavior. To generate the empirical city-size distribution, the ER economy has to amplify and capitalize on what little variance its degree distribution has to offer (cf. \text{proposition 2.3.1}). As a result \( \tau \) has to be ludicrously large to make things work. On the other hand, if the transportation infrastructure is in its early stage of development without any hubs, then the country’s transportation cost will probably be higher than more BA-like countries because Zipf’s law is a universally observed phenomenon. There is a trade-off between \( \tau \) and how close the transport network is to BA, provided that Zipf's law holds at all times.

\(^{18}\)The estimate tends to decrease as data size \( J \) increases.
of \cite{JR07} type. Regardless of how it came about, a network with a fat-tailed degree distribution results in the city-size distribution that closely resembles the actual distribution.

\section*{2.5 Conclusion and Extensions}

We examined how the network of cities affects the city-size distribution. We built a simple economic model with an explicit transport network. The bridge between network structure and city size is represented in \cite{18}, where we learned that there is a log-linear relationship between city size and city degree.

We put two commonly studied networks to the test. The classical ER random graph is too egalitarian to generate gravitationally large cities like New York City. The BA model explains the city-size distribution better than the ER model and bears very close comparison with other proposed city-size models in existence. The BA network has a scale-free degree distribution and the resulting city-size distribution behaves similarly via \cite{18}. In fact, it would be odd if the city-size distribution was not scale free under a BA network. Large nodes with a high degree like Chicago attract a large mass of people because A) goods produced in Chicago are in high demand for its inexpensive delivered price owing to its high degree and B) goods available for consumption in Chicago are also inexpensive thanks to its high degree. The exact opposite applies to small cities. But there are still some people knowingly living in small cities because we cannot afford to wipe them off the map due to preference for variety. This gives rise to a few cities of an overwhelming size and a myriad of small cities. The actual city-size distributions (we tried Belgium and the United States in particular) unanimously opt for a BA network.

From this point on, it would be reasonable to combine GEV to determine firm productivity as in \cite{BW14} and BA for transportation network structure by way of simulations, but we will not have an analytical solution due to the added complexity.

We argued that network structures motivate the population to form a specific distribution of city sizes. The structure of the network is pre-selected. Considering the fact that it is easier to
relocate people than to build transport infrastructure, this is not an unreasonable assumption in
the short run. New York City would have been much smaller had it not been the entrepôt to
Europe. However, the degree-city relationship is not a one-way street. It may be the other way
around: the relocation of people forces the transportation network to follow a specific pattern.
It can also be the case that the network structure and its associated city-size distribution are in
fact a product of some common underlying causes. The United States has seen a number of
drastic changes in its network structure. Tracing the historical co-development of the network
structure with the city-size distribution may reveal a clue to identifying the direction of causality.
A problem with this methodology is that transportation networks are not unique, in that there
are generally multiple modes of transport and multiple companies providing services in each
mode.

For now, as a preview, consider a commodity transportation firm that is installing a trans-
portation network that maximizes its profit by choosing \( r \). It uses the mechanism described by
\[JR07\] to add links to its network, namely random links and friends of friends, where \( r \) de-
termines the relative proportions. As noted in the introduction, the difference between social
networks and transportation networks is who makes the decisions about links, the nodes them-
selves or the owner of the network. If the revenues and cost of the network are additive across
nodes, then the profit from a network is additive across nodes, so there is no distinction between
maximizing the objective of a node and maximizing profit or utility of the entire network. In
other words, profit of the network owner corresponds to efficiency of the network in \[JR07\].
Suppose that shipping industry is competitive and the shipping firm’s indirect revenue function
is additively separable across nodes and convex in node degree, and also assume that its cost
is additive across nodes and proportional to node degree. Then we can follow the framework
proposed in section IV of \[JR07\], in particular Corollary 1, to find \( r \) that maximizes its profit, in-
dependent of the city-size distribution. This allows us to take the network development process
as exogenous, and leading to the BA network \( r = 0 \).

We finish our discussion with one last remark. It has been suggested that other networks be
implemented in our framework, for example the optimal transport network for a given population distribution (assuming a cost function). This would require the geodesic length or degree distribution for the optimal network. We are not aware of any results addressing this issue.
Chapter 3

A Spatial Production Economy Explains Gross Metropolitan Product

3.1 Introduction

Four out of five people live in cities, and they do so for various reasons, i.e., better job prospects, decent wage, urban amenities, or family obligations. The resulting size distribution of cities has kept the rapt attention of urban economists, and we now have a growing understanding of what it is and how it came about; however, the story does not end there. No one moves in or out of a city just for the sake of making its size larger or smaller, nor does the city size itself feed its population. The overriding research objective in the literature is the welfare implication of the city-size distribution, but the empirical distribution of GMP has never been analyzed to this date. We will take one step forward to show that the GMP distribution follows a fat-tail distribution and provide a theoretical background behind the relationship between city size and corresponding GMP.

Our major findings are as follows. First, two empirical regularities on the city-size distribution carry over to GMP. Most of GDP are generated in only a few cities just as the city-size distribution, the regularity known as Zipf’s law ([Gab99b], [Dur07]). In fact only 20% of cities create as much as 78.75% of urban GDP.1 GMP has a lower Pareto coefficient than city-size counterpart, i.e., its tail end is even heavier than the city-size distribution. Gibrat’s law also extends to GMP, as urban economic growth rates are independent of its GMP size. Second, GMP exhibits increasing returns to employment. That is, New York’s GMP is larger than any other city’s, even size for size. This is consistent with our first finding that the GMP distribution has a heavier tail than the

---

1 This relation is known as a 20-80 rule: 20% of agents are accountable for 80% of the results, a typical sign that something of scale-free nature is at work.
city-size distribution. We build a production economy model and establish that agglomeration economies are due to the trade-off between externalities and housing consumption. We prove that the equilibrium city size has to be such that an additional resident will reduce a housing lot size in the city but make up for it by raising citywide productivity.

In the existing city-size models, with the assumption of free mobility, consumers/workers will update their locations until they exhaust the locational arbitrage opportunities. Thus, regardless of the city size, cities become indifferent to consumers in equilibrium. This does not imply that workers are equally productive or their income will be the same across the board. In practice, per capita GMP varies by location (cf. figure 19 and table 7). People enjoy the same utility level at the end of the day\(^2\) but what induces interurban migration depends on GMP. People relocate to a city not for the sake of its size alone but for what its size has to offer, one of which is its GMP. The city-size distribution is the result of interurban migration. We will reveal the distribution of GMP and decode the economic forces behind it in this paper.

Figure 20 (in color) is a map of the United States with metropolitan statistical areas (MSA) colored according to their population density and GMP in 2010.\(^3\) Figure 20(a) comes with no surprise. It is well documented that the city-size distribution is tail heavy. What is newsworthy is figure 20(b)\(^4\). GMP shares the same pattern to city size in terms of distribution. Figure 21 represents the probability density function (PDF) and rank-size plot of GMP in 2010.

New York accounts for the lion’s share of GDP, followed by Los Angeles, and there are lots

---

\(^2\) Cities are put in equilibrium either by equating wage (e.g., [Dur07]) or utility level (e.g., [Gab09b]).

\(^3\) Population data also include micropolitan statistical area along with MSA. For definition of MSA, see [http://www.census.gov/population/metro/about/](http://www.census.gov/population/metro/about/).

\(^4\) Data source: Bureau of Economic Analysis.
(a) Population Density in 2010 (persons/km²).

(b) GMP in 2010 (in 2005 USD).

**Figure 20.**

(a) PDF of GMP. Dots are size proportionate to GMP.

(b) PDF of GMP in log scale.

**Figure 21.** PDF plots of GMP. See table 7 for the explanation of the selected cities above.

of mid-sized cities that are dwarfed by the high-ranked cities.

Our intended contribution is to provide a systematic understanding of the distribution of GMP. The conventional range of study on GMP has been limited within a city. Usual questions are in lines of how to promote the urban growth in Detroit, or the effect of overproduced liquid
natural gas in Pittsburgh. These fact-finding works and analyses of local economies play a part in the GMP distribution. GMP distribution is, after all, the accumulation of all these local economic activities combined. On the other hand, we found a holistic approach to the GMP distribution missing in the literature: GMP is reported in each city; therefore there will be a GMP distribution. We would like to get an aerial view from coast to coast and address GMP from the general equilibrium perspective.

There are two lines of research related to our project: one on the distribution of city sizes and the other on agglomeration economies. The first line of research studies the distribution of city sizes but not GMP, while the second one studies GMP but not the distribution thereof. We will fill in the gaps in this paper.

Overshadowed by the consuming interest in the city-size distribution, research into the distribution of GMP is nonexistent. Those who quote the citywide production function use it as an intermediate step to reach the equilibrium city-size distribution. As a byproduct, we get the equilibrium production level in each city, but predicted GMP has never been tested with any empirical data. Their primary objective is to explain the city-size distribution. We will take the GMP distribution as a byword rather than a byproduct.

On the other hand, the second line of work homes in on the question of how much of a boost we get by producing goods and services in a crowd rather than in a rural setting. The question is imperative because if there is no scale economies in cities, then there is no convincing reason to reside in a large but crowded city, barring other centripetal forces such as local public goods or access to a large, diversified labor pool [ABL07]. A study on citywide productivity becomes an essential part of the examination on city size (cf. [Hen74], [KKSS05]). In fact increasing returns to scale is one of the main ingredients in the formation of a city (Krugman [Kru91]). See Moomaw [Moo83] for review of earlier work in this literature.

Despite having related research agenda, these two lines of work take different approaches to theorizing about their respective target objective. The city-size distribution models based on

\footnote{But not necessary. Cf. [BK00].}
general equilibrium typically do not include capital stock as part of production function ([Duro07], [Eeco04] for example),\(^6\) whereas most agglomeration models do. Labor alone serves its purpose to explain the actual city-size distribution without involvement of capital stock. We sided with the city-size distribution models for our purposes. It is easy to measure a city size, but measuring citywide capital stock is not as straightforward as a head count. In fact there are no data on the level of capital stock at city level in the United States. Those studies that quote capital stock use the estimated level based on factors related to capital such as local public goods, housing and state roads, mixed in with predetermined weights ([Seg76]), or estimated retrospectively from the pair of labor and GDP per capita at city level ([Sve75]). Capital stock is known to be correlated with city size, which causes a multi-collinearity problem. According to [Seg76], capital stock’s contribution to GMP is .116 as opposed to labor’s .891. We did not test our model on capital stock but it is general enough to incorporate investment if need be.

The data set we use is more inclusive than any previous studies. There are 366 cities with accompanying GMP figures. The largest sample size used so far to test GMP is 30 by Mion and Naticchioni [MN05] according to [MGN09].\(^7\) For example cities like Beaumont, TX are too small to be included in the data set in [Seg76]. At the time of writing 366 is the largest data size for which GMP is reported.

GMP data also mesh with city-size data to provide an added layer of empirical validation to the existing models on the city-size distribution. The models of city-size distribution are empirically tested on the basis of city size, and the choice of city size as a data set to pitch against a model is obvious because they are built to explain the city-size distribution after all; however, they also need to be crosschecked with other spatial data, including rent or wage in each city. Otherwise a model can only explain city-size distribution but nothing else, which undermines its legitimacy as an urban economic model. GMP is one of those spatial data that complements city-size data to confirm a model’s relevance to the reality.

---

\(^6\) There are some exceptions. For example, Rossi-Hansberg and Wright [RHW07] address city-size distribution with capital stock incorporated into the model. Even then, actual capital stock level is not used for empirical testing.

\(^7\) The aforementioned study [Seg76] has 58 locations but output is limited to the manufacturing sector rather than GMP as a whole. These studies often quote census for manufacturers alone.
The remainder of the paper is organized as follows: section 3.2 investigates into the nature of GMP distribution and provides descriptive statistics on GMP along with city size. In section 3.3 we introduce the spatial production economy model to explain the findings in section 3.2 before we empirically evaluate our model’s performance in section 3.4. Section 3.5 concludes our study.

3.2 GMP Actualities

We will establish the Zipf’s and Gibrat’s law for GMP and also identify the relationship between GMP and city size. The US Bureau of Economic Analysis reports annual GDP by MSA along with the US GDP and estimated employment. Descriptive statistics for the employed data are in table 7.
<table>
<thead>
<tr>
<th>Data</th>
<th>Employment</th>
<th>GMP</th>
<th>GMP per capita</th>
<th>GMP growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit</td>
<td>MSA</td>
<td>Million 2005 USD</td>
<td>2005 USD</td>
<td>Annual rate</td>
</tr>
<tr>
<td>Year</td>
<td>2010</td>
<td>2010</td>
<td>2010</td>
<td>2010 over 2005</td>
</tr>
<tr>
<td>Coverage (MSA/USA)</td>
<td>84.90%</td>
<td>87.60%</td>
<td>87.60%</td>
<td>87.60%</td>
</tr>
<tr>
<td>Largest city</td>
<td>#001 New York</td>
<td>#001 New York</td>
<td>#001 Midland, TX</td>
<td>#001 Midland, TX</td>
</tr>
<tr>
<td>Largest size</td>
<td>18,919,787</td>
<td>1,147,160</td>
<td>89,350</td>
<td>11.60%</td>
</tr>
<tr>
<td>73rd largest city</td>
<td>#073 Akron, OH</td>
<td>#073 Worcester, MA</td>
<td>#073 Waterloo, IA</td>
<td>#073 Dallas, TX</td>
</tr>
<tr>
<td>Size share of #1-73</td>
<td>71.80%</td>
<td>78.75%</td>
<td>29.40%</td>
<td>N/A</td>
</tr>
<tr>
<td>City near arithmetic mean</td>
<td>#072 North Port, FL</td>
<td>#063 Madison, WI</td>
<td>#150 Greenville, SC</td>
<td>#171 Topeka, KS</td>
</tr>
<tr>
<td>Arithmetic mean</td>
<td>707,308</td>
<td>31,763</td>
<td>36,002</td>
<td>.49%</td>
</tr>
<tr>
<td>Median city</td>
<td>#183 Laredo, TX</td>
<td>#183 Bellingham, WA</td>
<td>#183 Gainesville, FL</td>
<td>#183 Monroe, LA</td>
</tr>
<tr>
<td>Median size</td>
<td>251,539</td>
<td>8,414</td>
<td>34,048</td>
<td>.35%</td>
</tr>
<tr>
<td>Smallest city</td>
<td>#366 Carson City, NV</td>
<td>#366 Palm Coast, FL</td>
<td>#366 Palm Coast, FL</td>
<td>#366 Lake Charles, LA</td>
</tr>
<tr>
<td>Smallest size</td>
<td>55,212</td>
<td>1,132</td>
<td>11,793</td>
<td>-5.98%</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>1,582,442</td>
<td>86,824</td>
<td>11,257</td>
<td>.0212</td>
</tr>
<tr>
<td>Skewness</td>
<td>6.669</td>
<td>7.500</td>
<td>1.409</td>
<td>.9088</td>
</tr>
<tr>
<td>City near geometric mean</td>
<td>#149 Naples, FL</td>
<td>#154 Kalamazoo, MI</td>
<td>#178 Mobile, AL</td>
<td>N/A</td>
</tr>
<tr>
<td>Geometric mean</td>
<td>321,416</td>
<td>11,078</td>
<td>34,460</td>
<td>N/A</td>
</tr>
<tr>
<td>Mean of log value</td>
<td>12.68</td>
<td>23.13</td>
<td>10.45</td>
<td>N/A</td>
</tr>
<tr>
<td>Standard deviation of log value</td>
<td>1.062</td>
<td>1.216</td>
<td>.2920</td>
<td>N/A</td>
</tr>
<tr>
<td>Skewness of log value</td>
<td>1.109</td>
<td>1.103</td>
<td>.2251</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Table 7. Descriptive Statistics. The statistics above the line (shaded in blue) are related to a linear scale and below the line (shaded in green) are related to a log scale. The mean of log value is same as the log of geometric mean. The first 73 cities make up for the upper 20% of the total number of cities.
3.2.1 Gibrat’s Law for GMP

For starters, we looked into the dynamics: Does a large GMP make a city grow fast? The answer: no. Gibrat’s law implies that the size of a city does not have any bearing on its growth rate. The city-size distribution is known to follow Gibrat’s law well ([IOo3]). It turns out that GMP does the same. We carried out both parametric and non-parametric estimations following [Eeco4] to examine the relationship between GMP and GMP growth rate.

3.2.1.1 Non-Parametric Estimation

First we estimate the conditional expectation of GMP growth rate \( E[g|Y] = m(Y) \) with a Nadaraya-Watson kernel estimator [Wat64]

\[
nm(Y) = \sum_{i=1}^{I} \frac{g_i K_h(Y - Y_i)}{\sum_{j=1}^{I} K_h(Y - Y_j)}.
\]

\( Y \) denotes GMP and \( g \) denotes its growth rate. Sample size is \( I = 366 \) with each city indexed by a superscript \( i \). \( K_h(\cdot) \) is a scaled kernel with a bandwidth \( h \). We gathered data from 2005 and 2010 to compute growth rates. For non-parametric estimation we standardize the growth rate to take out the nationwide growth rate.\(^8\) GMP is defined by the geometric mean \( Y := \sqrt{Y_{05}Y_{10}} \), assuming exponential growth. Figure 22 plots the growth rate and its kernel estimation. We tried to estimate \( nm(Y) \) first (figure 22(a))\(^9\). The disperse spread of GMP towards the upper end swings the estimate from side to side and makes it hard to interpret the relationship. We went for a log of GMP instead and recorded the result in figure 22(b), which now exhibits a discernible pattern. There seems to be a slight inclination to the left and right tails, probably because of a smaller number of observations to the both ends than in the rest of the range. Other than that, our estimate seems to be in support of the Gibrat’s law for GMP. For analysis of variance, see

---

\(^8\) In particular we take the difference between \( log(Y_{10}) - log(Y_{05}) \) and the sample mean, divided by the standard deviation to be the normalized growth rate \( g \).

\(^9\) We had to stretch the bandwidth further than the usual width of 2.727e+10 to cover up the large gap between New York and Los Angeles. The estimated growth rate is positive for GMP larger than 9e+11 and above solely because of New York and significantly lower than zero because of Los Angeles around 6e+11 to 9e+11.
appendix A.7.

3.2.1.2 Parametric Estimation

Next, we regress GMP growth rate on GMP. Estimates are reported in table 8. We first regressed GMP growth rate on GMP. Figure 23(a) seems to indicate that the regression line is pulled upwards partly because of New York. To counteract this sensitivity to large cities, we regressed GMP on the log of GMP as well (figure 23(b)).

<table>
<thead>
<tr>
<th>Regressor</th>
<th>Coefficient</th>
<th>t-statistic</th>
<th>R²</th>
<th>Figure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4.482e-03</td>
<td>3.84</td>
<td>8.077e-04</td>
<td>23(a)</td>
</tr>
<tr>
<td>GMP</td>
<td>6.981e-15</td>
<td>.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(GMP)</td>
<td>23(b)</td>
<td></td>
<td>1.230e-03</td>
<td>5.044e-03</td>
</tr>
<tr>
<td>t-statistic</td>
<td>-2.373e-02</td>
<td>1.36</td>
<td>1.230e-03</td>
<td>5.044e-03</td>
</tr>
</tbody>
</table>

Table 8. Ordinary least squared (OLS) estimate of growth rate.

---

10 We did not standardize the GMP growth rate for parametric estimation. The intercept will capture the nationwide growth rate.

11 The coefficient on GMP may well have been negative had New York’s growth rate been negative. The estimates’ dependence on New York is not all that welcoming because, while it is large, New York is still just one observation as much as Beaumont, TX is.
The null is not rejected at the 5% level of confidence on GMP or on the log thereof. Once again the estimates seem to agree with the Gibrat’s law.

![Graph showing GMP growth rate vs. geometric mean GMP](image)

**Figure 23.**

### 3.2.2 Zipf’s Law for GMP

As we have seen in figure 21, GMP seems to be well traced by a power law. OLS estimation confirms the power-law behavior of GMP, as documented in table 9 and figure 24. The Pareto

<table>
<thead>
<tr>
<th>Regressor</th>
<th>Coefficient</th>
<th>t-statistic</th>
<th>$R^2$</th>
<th>Figure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>16.34</td>
<td>-174.57</td>
<td>.9763</td>
<td>24(a)</td>
</tr>
<tr>
<td>log(Employment)</td>
<td>-9003</td>
<td>-122.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(GMP)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coefficient</td>
<td>23.13</td>
<td>-152.89</td>
<td>.9756</td>
<td>24(b)</td>
</tr>
<tr>
<td>t-statistic</td>
<td>-7875</td>
<td>-120.58</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 9.** Rank-size and rank-GMP regression

exponent is -.9003 on employment\(^{12}\) whereas we have -.7878 on GMP. This is indicative of the fact that the GMP distribution is even more skewed than the corresponding city-size distribution. This is to be theoretically verified with proposition 3.3.2. It should be noted that the use of OLS

\(^{12}\)Employment data are based on population estimates that the Bureau of Economic Analysis uses to compute per capita GMP.
for rank-size plot is now obsolete. We have included OLS just for reference. As pointed out by Gabaix and Ioannides [GI04], the city-size distribution does not sit well with the assumptions on errors in OLS estimation. The same criticism applies to the GMP distribution as well. Also, due to the limited data range, it is likely that Zipf’s law applies only to the upper tier and that the un-truncated GMP distribution has a fat tail to the left as well (cf. [Eco04]). In this case, a distribution other than a Pareto distribution, such as a lognormal or double Pareto lognormal ([GZS10]), may be an apt choice to describe the data. We will reconfirm Zipf’s law for GMP both theoretically (in section 3.3.2.1) and empirically (in section 3.3.3) without OLS. The case in point is not whether Zipf’s law describes the upper end of the distribution in particular but that the GMP distribution has a fat tail.

### 3.2.3 City Size and GMP

Figure 25 shows the relationship between working population and the aggregate product in a city. There seems to be a log-linear relationship between them with coefficient slightly but statistically significantly larger than one, indicating increasing returns to scale between city size and GMP.
Table 10 reports the results with figure 25.\textsuperscript{13}

The numbers are not too far off from the findings from the second line of work mentioned in section 3.1. For example Shefer [She73] finds that a 1\% rise in input will results in a 1.12\% increase in output (note; however, that this is just for the primary metal industry, whereas our numbers are for GMP).

\textsuperscript{13}Note that
\[
\log(Y/L) = \gamma_0 + \gamma_1 \log L \quad \Rightarrow \quad \log Y = \gamma_0 + (\gamma_1 + 1) \log L
\]
on a per-capita basis. On aggregate level, \( \log Y = \beta_0 + \beta_1 \log L \) so that \( \gamma_1 = \beta_1 - 1 \), as can be seen in table 10.
Figure 25.
<table>
<thead>
<tr>
<th>Regressand</th>
<th>Parameter</th>
<th>Value</th>
<th>Intercept</th>
<th>Employment</th>
<th>log(Emp.)</th>
<th>$\bar{R}^2$</th>
<th>$R^2$</th>
<th>$\bar{R}^2$</th>
<th>Fig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMP</td>
<td>Actual</td>
<td>Coefficient</td>
<td>-6.492e+09</td>
<td>5.409e+04</td>
<td>111.81</td>
<td>.9717</td>
<td>.9716</td>
<td>25(a)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t$-value</td>
<td>-7.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(GMP)</td>
<td>Actual</td>
<td>Coefficient</td>
<td>8.960</td>
<td></td>
<td></td>
<td>1.117</td>
<td>.9528</td>
<td>.9527</td>
<td>25(b)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t$-value</td>
<td>54.02</td>
<td></td>
<td></td>
<td>85.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(GMP)</td>
<td>Theoretical</td>
<td>Coefficient</td>
<td>8.165</td>
<td></td>
<td></td>
<td>1.180</td>
<td>.9498</td>
<td>.9497</td>
<td>25(b)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t$-value</td>
<td>47.74</td>
<td></td>
<td></td>
<td>87.79</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GMP per capita</td>
<td>Actual</td>
<td>Coefficient</td>
<td>3.422e+04</td>
<td>2.522e-03</td>
<td></td>
<td>.1257</td>
<td>.1233</td>
<td>25(c)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t$-value</td>
<td>56.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(GMP per capita)</td>
<td>Actual</td>
<td>Coefficient</td>
<td>8.960</td>
<td></td>
<td></td>
<td>.1173</td>
<td>.1821</td>
<td>.1799</td>
<td>25(d)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t$-value</td>
<td>54.02</td>
<td></td>
<td></td>
<td>9.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(GMP per capita)</td>
<td>Theoretical</td>
<td>Coefficient</td>
<td>8.165</td>
<td></td>
<td></td>
<td>.1800</td>
<td>.1302</td>
<td>.1278</td>
<td>25(d)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t$-value</td>
<td>47.74</td>
<td></td>
<td></td>
<td>13.39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Housing</td>
<td>Actual</td>
<td>Coefficient</td>
<td>-1.736e+09</td>
<td>8489</td>
<td></td>
<td>.9571</td>
<td>.9569</td>
<td>25(e)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t$-value</td>
<td>-9.88</td>
<td></td>
<td></td>
<td>84.60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(Housing)</td>
<td>Actual</td>
<td>Coefficient</td>
<td>4.149</td>
<td></td>
<td></td>
<td>1.305</td>
<td>.8769</td>
<td>.8765</td>
<td>25(f)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t$-value</td>
<td>11.98</td>
<td></td>
<td></td>
<td>47.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(Housing)</td>
<td>Theoretical</td>
<td>Coefficient</td>
<td>5.732</td>
<td></td>
<td></td>
<td>1.180</td>
<td>.8688</td>
<td>.8684</td>
<td>25(f)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t$-value</td>
<td>16.03</td>
<td></td>
<td></td>
<td>41.88</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 10. $\bar{R}^2$ is an adjusted value of $R^2$. For $t$-value, the null is coefficient equals zero, whereas for $t$-statistic, the null is coefficient equals theoretical value. Section 3.4.2 explains theoretical value.

### 3.3 Model

#### 3.3.1 Spatial Production Economy

We construct an intercity general equilibrium model to seek a comprehensive explanation for all the empirical findings in section 3.2. In particular we develop a production economy with three commodities: composite goods, housing and leisure,\(^{14}\) and two types of agents: worker/consumer and landlord.

There are $I$ cities in the economy. $S^i$ residents live in city $i$, totalling $S = \sum_{i=1}^{I} S^i$ of urban population nationwide. Each city has a demographic similar to the Alonso model (cf. Berliant and Fujita [BF92]). See figure 26 for one example representation of agents involved in this production

\(^{14}\)Alternatively, we can include capital goods but due to lack of data, we limit ourselves to three goods in this economy.
economy. Each city has a landlady who owns all the area \( H \) in city \( i \). She is retired and lives off her rental income \( r_i H \), where \( r_i \) marks the city’s rental rate (think of her as the first settler in town or a developer). She is an immobile\(^{15} \) landlady and assumed to consume only composite goods and leisure out of her one unit of allotted time.\(^{16} \) The remainder of the urban population are mobile, active and identical workers/consumers who supply labor \( l_R^i \) out of their one unit of allotted time to produce a basket of goods \( c_R^i \) that includes all the goods and services other than housing \( h_R^i \) and leisure \((1 - l_R^i)\). Their consumption bundle \( x^i \) and endowment \( e^i \) are given by

\[
x_R^i = \begin{pmatrix} c_R^i \\ h_R^i \\ 1 - l_R^i \end{pmatrix}, \quad x_L^i = \begin{pmatrix} c_L^i \\ h_L^i \\ 1 - l_L^i \end{pmatrix}, \quad e_R^i = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad e_L^i = \begin{pmatrix} H \\ 1 \end{pmatrix},
\]

where subscript \( R \) denotes a representative working Resident and \( L \) denotes the Landlady.

On the production side, there are many firms in the city who employ one worker each and produce the identical immobile commodity in a perfectly competitive environment. The production plan \( y^i \) of a representative firm is given by

\[
y^i = \begin{pmatrix} f(l_F^i, h_F^i; S^i) \\ -h_F^i \\ -l_F^i \end{pmatrix},
\]

where \( l_F^i \) denotes labor demand and \( h_F^i \) denotes land input used as a production site. We let the production function \( f(\cdot) \) depend on the city size to allow for externalities within the city such as knowledge spillover effects or congestion to have an impact on productivity of individual firms.

\(^{15} \) We assume that she cannot change her city of residence so that we can count the rental income toward GMP where it is collected. Otherwise the rental income may be included in the city where she actually lives, which may not be the city whose land she owns if she is an absentee landlady. However, we will not count her toward \( S^i \) for notational ease. We will return to the role of her location choice in section 3.4.2.

\(^{16} \) Assume that she lives in the city where she is a landlady but in the special lot designated for her outside \( H \) to keep our analysis tractable. She needs to live in the city where she is a landlady because of \textit{footnote 15}. 

80
in the same city.\textsuperscript{17}

The intracity production economy in city \(i\) is identified by

\[
\mathcal{P}^i := \left\{ \left( X^i_{N^i}, e^i_N \right) \in \mathbb{R}_{\geq 0}^n, Y^i_F \right\},
\]

where \(X^i_{N^i} := \mathbb{R}_+^3\) is a consumption set of a representative worker or the landlady, \(\geq_N\) is a complete preorder over consumption set \(X^i_{N^i}\), and \(Y^i_F\) is a production set of a representative firm given by

\[
Y^i_F := \left\{ y' = (c^i_F, h^i_F, l^i_F) \in \mathbb{R}_+^3 : c^i_F \leq f(l^i_F, h^i_F, S^i) \right\}.
\]

A feasible allocation in \(\mathcal{P}^i\) is defined as follows:

**Definition 3.3.1: Feasible Allocation**

For given \(S^i \in [0, S]\), an allocation \((x^i_{R^i}, x^i_{L^i}, y') \in X^i_{R^i} \times X^i_{L^i} \times Y^i_F\) in the intracity production economy \(\mathcal{P}^i\) is feasible iff

\[
x^i_{R^i}S^i + x^i_{L^i} = y'S^i + e^i_{R^i}S^i + e^i_{L^i}.
\]

To find GMP we need to compute the value of each commodity. Composite goods are measured in a number of baskets, housing in ft\(^2\) and leisure in hours for example. We cannot add them together without weighing them with the price to derive a sensible measure of a city’s production and earnings. Let \(p^i := (1, r^i, w^i)'\) be the price on a composite good, lot size and leisure. We take composite goods as a numéraire.\textsuperscript{18} There are two equivalent ways to define GMP. From the production point of view, GMP \(\mathcal{Y}^i\) is defined by the total value of all the final goods and services produced in the city, \(\mathcal{Y}^i = p^i \cdot (y' + e^i_{R^i}S^i + e^i_{L^i})\). From the consumers’ end, GMP is the sum of all the expenditures on goods and services, \(\mathcal{Y}^i = p^i \cdot (x^i_{R^i}S^i + x^i_{L^i})\). They come out to the same number due to Walras’ law.

**Definition 3.3.2: GMP**

\textsuperscript{17} Since we bundle all the goods in a single basket, there is no distinction between localization economies (agglomeration economies within an industry) and urbanization economies (agglomeration economies across the industries within a city) in our model.

\textsuperscript{18} Note that none of the commodities are tradable beyond the city border in this economy.
GMP in the intercity production economy $\mathcal{P}^i$ of size $S^i$ is identified by

$$\mathcal{Y}^i := p^i \cdot (y^i + e^i_R S^i + e^i_L) = p^i \cdot (x^i_R S^i + x^i_L).$$  \hspace{1cm} (24)$$

In application GDP does not count leisure time. We consume leisure for the price of the opportunity cost (namely, lost wage), but in practice there is no explicit/accounting trace of market transactions for the consumption of leisure to track down the leisure portion of GDP. In particular we produce and consume

$$w^i \left\{ \left( 1 - l^i_R \right) S^i + \left( 1 - l^i_L \right) \right\} \text{ worth of leisure, but this part is excluded from recorded GDP and by extension, from GMP as well.}$$

We use the data provided by the US Bureau of Economic Analysis, and their figures are based on tax reports. We do not pay tax on leisure consumption. Thus, we shall redefine $\mathcal{Y}^i$ with only the first two entries and take out the last entry (leisure)

$$\begin{bmatrix} 1 \cr r^i \end{bmatrix} \cdot \begin{bmatrix} c^i_R \cr c^i_L \cr h^i_R \cr h^i_L \end{bmatrix} \cdot S^i + \begin{bmatrix} f(l^i_F, h^i_F; S^i) \cr h^i_F \end{bmatrix} \cdot S^i + \begin{bmatrix} 0 \cr S^i \cr 0 \cr H \end{bmatrix}$$

for statistical purposes.

As we understand from our empirical findings in section 3.2, $\mathcal{Y}^i$ exhibits increasing returns to scale $S^i$. Then according to (25), at least one of $f(\cdot)$ or $r^i$ (possibly both) needs to be increasing more than proportionately in $S^i$ to explain the actual GMP distribution as a function of the city size.

To find the equilibrium price vector, first define $\theta^i := (\theta^i_R, \theta^i_L)$ as a vector of a representative resident and landlady’s share of profit ($\theta^i_R, \theta^i_L \in [0, 1]$ and $\theta^i_R S^i + \theta^i_L = 1$).

**Definition 3.3.3: Intracity Equilibrium**

For a given $\theta^i$ and $e^i$, an intracity equilibrium in city $i$ is a feasible allocation $(x^i_R, x^i_L, y^i)$ and price vector $p^i$ such that

1. For $N = R$ and $L$

$$p^i \cdot x^i_N \leq p^i \cdot e^i_N + \theta^i_N p^i \cdot y^i S^i.$$  \hspace{1cm} (26)$$
2. For \( N = R \) and \( L \)

\[
p^i^* \cdot x^i_N \leq p^i^* \cdot e^i_N + \Theta^i_N p^i^* \cdot y^i^* S^i_i \quad \Rightarrow \quad x^i^* \geq_N x^i,
\]

for any \( x^i_N \in X^i \).

3. For any \( y_i \in Y^i_F \),

\[
p^i^* \cdot y^i^* \geq p^i^* \cdot y^i.
\]

To identify the equilibrium city size, let the intercity production economy \( \mathcal{P} := \left\{ (\mathcal{F}^i, S^i_i)^I = 1 \right\} \) and define

**Definition 3.3.4: Intercity Equilibrium**

For a given ownership matrix \( (\theta^i_i)_{i=1}^I \in [0, 1]^I \) and endowment matrix \( (e^i_i)_{i=1}^I \in \prod_i (X^i_R \times X^i_L) \), an intercity equilibrium in the production economy \( \mathcal{P} \) is a list of a feasible allocation matrix \( (x^i_R^*, x^i_L^*, y^i^*)^I_{i=1} \in \prod_i (X^i_R \times X^i_L \times Y^i_F) \), price matrix \( (p^i^*)^I_{i=1} \in \mathbb{R}^3_I \), and size distribution \( (S^i)_{i=1}^I \in [0, S]^I \) such that for any \( S^i > 0 \),

1. \( (x^i_R^*, x^i_L^*, y^i^*, p^i^*) \) is an intracity equilibrium for any city \( i \).
2. For any \( i, j \),

\[
x^j_R^* \sim_R x^i_R^*.
\]
3. Urban population adds up to

\[
\sum_i S^i = S.
\]

The second item (29) is due to free mobility of workers. This does not apply to landladies, who are locked in their place of residence to keep the housing portion of GMP where it is generated.

The equilibrium city-size distribution is the size component of an equilibrium in \( \mathcal{P} \) and the GMP distribution is (25) computed with an equilibrium in \( \mathcal{P} \).
3.3.2 Application

To derive the exact distribution of GMP for empirical testing, consider an application of the spatial production model developed in section 3.3.1 with production function and labor market in the style of Eeckhout [Eec04] with the explicit presence of landladies (see figure 26 for a schematic representation of the agents and commodities involved in this example). We will find the analytical solution to the intercity equilibrium, from which we obtain the equilibrium GMP distribution.

3.3.2.1 Intracity Equilibrium

To start off, pick any city $i$ and consider its intracity equilibrium. Firm’s production plan is specified by

$$f\left(l_i, h_i; S^i\right) = A_i a_+\left(S^i\right) a_-\left(S^i\right) l_i,$$

(31)

where $A^i$ is a stochastic citywide productivity parameter, $a_+() (>0)$ measures the positive externality shared among the firms operating within the same city, and $a_-() (\in (0, 1))$ measures congestion externality. City size is assumed to raise the productivity of all the firms operating in the city. Positive externality enhances with size ($a'_+() > 0$). Each consumer supplies $l_k$ units of gross labor but congestion externality adversely affects effective labor. The fraction $1 - a_-\left(S^i\right)$ of labor will be spent on commuting rather than on production. The level of reduction in effective labor aggravates with the size of a city ($a'_-() < 0$). Firms do not pay for the time lost in commuting and workers assume responsibility for the time cost of commuting.

Figure 26. Commodity flow. Leisure is excluded in accordance with the practical definition of GDP adopted by the US Bureau of Economic Analysis.
That is, firms will pay (ostensible) wages at the rate of $\omega^i$ only for the fraction of $l^i_R$ when their worker is present at work, i.e., only for $a_-\left(S^i\right)l^i_R$ hours out of $l^i_R$. On an hourly basis, (effective) wage is knocked down to $w^i\left(S^i\right) := \omega^ia_-\left(S^i\right)$ for each hour devoted for work, inclusive of commuting time.\footnote{Hence, the opportunity cost of leisure is $w^\lambda$ rather than $\omega^i$.} We will discuss the role of a landlady’s labor supply later. We assume that firms do not require land as input in accordance with [Eco4] for simplicity, but land can readily be incorporated into our production economy in section 3.3.1 as a factor of production.

Profit (28) turns into

$$p^i \cdot y^i = A^i a_+ \left(S^i\right) a_- \left(S^i\right) l^i_F - \omega^i a_- \left(S^i\right) l^i_F = \left[B^i \left(S^i\right) - \omega^i a_- \left(S^i\right)\right] l^i_F,$$

where $B^i(S^i) := A^i a_+ \left(S^i\right) a_- \left(S^i\right)$. Since production function (31) exhibits constant returns to scale in $l^i_F$,

$$p^i \cdot y^i = 0$$

in equilibrium (otherwise $y^i$ violates profit maximiation condition (28)).\footnote{Constant returns to scale implies $y^\lambda \in Y^\lambda \Rightarrow \lambda y^\lambda \in Y^\lambda$ for any $\lambda \geq 0$. If $p^i \cdot y^\lambda > 0$, profit still improves with $2y^\lambda \in Y^\lambda$. If $p^i \cdot y^\lambda < 0$, profit still improves with $0y^\lambda \in Y^\lambda$.} Hence, if $l^i_F > 0$, it must follow that

$$B^i(S^i) = \omega^i a_- \left(S^i\right) \left(= w^i \left(S^i\right)\right)$$

in equilibrium.

Note here that aggregate production may exhibit agglomerative economies due to positive externality $a_+ (\cdot)$, but internal scale economies are still absent because individual production function is linear in $l^i_F$. For more on a dialectic between increasing and constant returns to scale, see Rossi-Hansberg and Wright [RHW07].

Next order of business is the consumers. Represent $\succeq_R$ in $\mathcal{P}^i$ by

$$u_R \left(c^i_R, h^i_R, l^i_R\right) = \alpha \log c^i_R + \beta \log h^i_R + \gamma \log \left(1 - l^i_R\right),$$

(35)
where $\alpha, \beta, \gamma > 0$, and assume $\alpha + \beta + \gamma = 1$ for simplicity. According to the feasibility condition (23) household income is given by $p^i \cdot c^i_R + \theta^i_R \cdot p^i \cdot y^i S^i$. Since firms earn zero profit (33), household income simplifies to labor income $p^i \cdot c^i_R = w^i \cdot (S^i) \cdot 1$ alone, with which to buy composite goods $c^i_R$, housing $h^i_R$ and leisure $(1 - \ell^i_R)$ at the price of $p^i = (1, r^i, w^i)$. Marshallian demand is

$$x^i_R(p^i, w^i) = \begin{pmatrix} c^i_R(p^i, w^i) \\ h^i_R(p^i, w^i) \\ 1 - \ell^i_R(p^i, w^i) \end{pmatrix} = \begin{pmatrix} \alpha w^i(S^i) \\ \beta w^i(S^i) / r^i \\ \gamma \end{pmatrix} = \begin{pmatrix} \alpha B^i(S^i) \\ \beta B^i(S^i) / r^i \\ \gamma \end{pmatrix}$$

The second equality holds as a result of profit maximization (28) and its consequence (34). Labor supply $\ell^i_R$ of a typical household will be $1 - \gamma = \alpha + \beta$. Material balance (23) requires that $(1 - \ell^i_R) S^i + 1 - \ell^i_L = -I^i_F S^i + 1 \cdot S^i + 1$. Since utility maximization for the retired landlady (27) results in $\ell^i_L^* = 0$ in this economy (see (39) below), $\ell^i_R = \ell^i_F^*$, which furthermore implies that the equilibrium production plan will be

$$y^{i^*} = \begin{pmatrix} f(\ell^i_F, h^i_F, S^i) \\ 0 \\ -\ell^i_F \end{pmatrix} = \begin{pmatrix} (\alpha + \beta)B^i(S^i) \\ 0 \\ -(\alpha + \beta) \end{pmatrix}.$$  

(37)

Turning to the landlady, represent $\succeq_L$ in $P^i$ by

$$u_L(c^i_L, h^i_L, \ell^i_L) = c^i_L \mathbf{1}_{[\ell^i_L = 0]}(\ell^i_L),$$

where $\mathbf{1}_{[\ell^i_L = 0]}(\cdot)$ is an indicator function that takes the value of one when $\ell^i_L = 0$ and zero otherwise. Since she is retired, any hour of labor $\ell^i_L > 0$ will instantly push her utility level down to zero regardless of an increment in her utility level from an increased consumption of composite goods financed through her labor income.\(^{21}\) That is, her utility level is nonnegative over the plane $\ell^i_L = 0$.

---

\(^{21}\)Alternatively, we could model her as an active worker, which complicates our notations without much gain in insights.
in $\mathbb{R}^+_3$ and zero elsewhere. Once again, since the share of zero profit (33) earns her nothing, the budget constraint (26) implies that the landlady’s income is $p^i \cdot e^i_L = r^i H + w^i (S^i)$. Her Marshallian demand is

$$x^i_L (p^i, w^i) = \begin{pmatrix} e^i_L (p^i, w^i) \\ h^i_L (p^i, w^i) \\ 1 - \frac{p^i}{H_L (p^i, w^i)} \end{pmatrix} = \begin{pmatrix} r^i H \\ 0 \\ 1 \end{pmatrix}. \tag{39}$$

Then residential utility maximization (36), profit maximization (37) and landlady’s utility maximiation (39) rewrite feasibility condition (23) as

$$\begin{pmatrix} \alpha B^i \left( S^i_R \right) \\ \beta B^i \left( S^i_R / r^i \right) S^i + 0 \\ 1 - (\alpha + \beta) \end{pmatrix} = \begin{pmatrix} \frac{(\alpha + \beta) B^i \left( S^i \right)}{S^i} \\ 0 \\ -(\alpha + \beta) \end{pmatrix} \begin{pmatrix} 0 \\ S^i \end{pmatrix} + \begin{pmatrix} 0 \\ S^i + H \end{pmatrix}. \tag{40}$$

from which, along with the first order condition (34), we can find the equilibrium price vector in city $i$ as

$$p^i \ast = \begin{pmatrix} 1 \\ r^i \\ w^i \end{pmatrix} = \begin{pmatrix} 1 \\ \frac{\beta B^i \left( S^i \right) S^i}{S^i / H^i} \\ B^i \left( S^i \right) \end{pmatrix}. \tag{41}$$

(see figure 26 to track the commodity flow in equilibrium over $P^i$). Observe that the rent $r^i$ goes up if 1) city $i$ draws a good technological shock $A^i$, 2) positive externality $a_+ \left( S^i \right)$ intensifies, or 3) city $i$ becomes less crowded ($a_- \left( S^i \right)$). Likewise the leisure becomes expensive for the same reasons. Reasons 2) and 3) are triggered by urban growth, whereas reason 1) is independent of $S^i$.

It is worth pointing out that our economy makes a judicious use of limited labor. The working population is capped at $S$. The intracity equilibrium and more noticeably, the intercity equilibrium, allocate more people to a city with a good production environment and pull back labor from a city of low productivity. Indirect utility is increasing in $B^i \left( S^i \right)$ in equilibrium. Migration
dynamics are such that there is an inflow when \( B^i(S^i) \) is above the national average and vice versa if it is below. The economy has an auto-rerouting mechanism built into it if population allocation ever deviates from the equilibrium, by incentivizing people to move to a productive city.\(^{22}\)

With feasibility condition (40) and the equilibrium price (41) we obtain GMP (24) in equilibrium as follows:

\[
Y^i = p^* \cdot (x^R_{S^i} + x^L_{S^i}) \\
= (\alpha + \beta)B(S^i)S^i + \beta B(S^i)S^i + B(S^i)(\gamma S^i + 1) \\
\text{Reported portion of GMP } Y^i (25) \\
= (1 + \beta)B(S^i)S^i + B(S^i). 
\] (42)

On the second line in (42) are the value of composite goods, housing and leisure for each. Only the first two are included in the actual GMP.

Equilibrium GMP (42) leads to the following:

**Proposition 3.3.1: Citywide Scale Economies in an Intracity Economy**

*Consider the equilibrium in an intracity economy \( P^i \). The reported portion of GMP \( Y^i \) exhibits increasing returns to scale in size \( S^i \) iff \( B^i(S^i)S^i \) exhibits increasing returns to scale in \( S^i \).*

**Proof.** Apparent from (42). \( \square \)

In reference to section 3.2.3, observed data seem to suggest that positive externality does outstrip negative externality. However, as we will see in proposition 3.3.2, citywide scale economies in intercity equilibrium will be positive without assuming increasing returns to scale on \( B^i(S^i)S^i \).

### 3.3.2.2 Intercity Equilibrium

To find the intercity equilibrium in definition 3.3.4, rewrite indifference principle (29) in terms of a utility function (35) so that indirect utility \( u(x^R_{S^i}) = u(x^L_{S^i}) \) for any \( i \) and \( j \) with \( S^i, S^j > 0 \). This

\(^{22}\) The allocation of labor is still not efficient though, due to externalities.
leads to

\[ B^i \left( S^i \right) \left( S^i \right)^{\frac{\beta}{\alpha}} = B^i \left( S^i \right) \left( S^i \right)^{\frac{\beta}{\alpha}} =: K, \tag{43} \]

where \( K \) is a location-invariant constant. According to (42), GMP is

\[ Y^i = (1 + \beta) K \left( S^i \right)^{\frac{\beta}{\alpha} + 1} + K \left( S^i \right)^{\frac{\beta}{\alpha}}, \]

and the reported portion of GMP will be

\[ Y^i = (\alpha + 2\beta) K \left( S^i \right)^{\frac{\beta}{\alpha} + 1} \tag{44} \]

in the neighborhood of the equilibrium size, which breaks down into composite goods production/consumption \((\alpha + \beta) K \left( S^i \right)^{\frac{\beta}{\alpha} + 1}\) and housing production/consumption of \(\beta K \left( S^i \right)^{\frac{\beta}{\alpha} + 1}\). Now we have

**Proposition 3.3.2: Citywide Scale Economies in an Intercity Economy**

*If an intercity economy \( P \) is in equilibrium, the reported portion of GMP \( Y^i \) exhibits increasing returns to size in the neighborhood of equilibrium size \( S^i \).*

*Proof.* Immediate from (44). \( \square \)

In comparison to proposition 3.3.1 it is curious that we have to assume \( B^i \left( S^i \right) S^i \) to be increasing returns to scale only in \( P^i \) but not in \( P \). The short answer to this enigma is that free mobility puts the city size where scale economies are in effect. For illustrative purposes, assume that \( a_+ \left( S^i \right) a_- \left( S^i \right) \) takes the form \( \left( S^i \right)^{\delta \left( S^i \right)} \). Note that proposition 3.3.1 specifically requires \( \delta \left( S^i \right) > 0 \) in \( P^i \) but proposition 3.3.2 does not because perfect mobility will bring \( \delta \left( S^i \right) \) above zero anyway.

Now assume that \( P \) is in equilibrium. Suppose that in some city \( i \), size \( S^i \) rose by one (call this new resident Axel). In this case, housing consumption will be reduced in the city because residents have to make room for Axel’s house out of a fixed supply of \( H \), and he also exacerbates congestion in the city. However, since \( P \) is in equilibrium, reduction in utility level by curtailed housing consumption needs to be offset by either \( c^l \) or \( l^l \) in compliance with utility equalization (29). Since \( l^l \) is independent of size (i.e., leisure consumption does not and cannot accommodate the change to the city residents introduced by Axel), compensation must be made through in-
creased consumption in \( c_R^i \) alone. Then the question is: Can he produce enough composite goods to leave everyone in city \( i \) on the same indifference curve?

It is the answer to this question that makes \( P \) increasing returns to scale in size in equilibrium. The marginal rate of substitution between \( c_R^i \) and \( h_R^i \) is \( \frac{\beta_c^i}{\alpha h_R^i} \) baskets for each \( \text{ft}^2 \). Now, Axel carves \( \frac{\partial h_R^i}{\partial S_i} = \frac{h_R^i}{S_i} \) \( \text{ft}^2 \) from every resident’s lot in the city. Thus, each city resident needs to have \[
\left( \frac{\beta}{\alpha} \frac{c_R^i}{h_R^i} \right) \frac{h_R^i}{S_i} = \frac{\beta}{\alpha} \frac{1}{S_i} \]
more baskets to keep to the countrywide utility level (or else the current allocation will not be an equilibrium). Then the addition of Axel into the city needs to raise the individual production of composite goods as follows:

\[
\frac{\partial f(\cdot; S^i)}{\partial S_i} = \frac{\beta}{\alpha} f(\cdot; S^i) \frac{1}{S^i} \\
\Rightarrow \frac{\delta(S^i)}{A^i(S^i) \delta(S^i)^{-1}} = \frac{\beta}{\alpha} A^i(S^i) \delta(S^i)^{\delta(S^i)-1} \\
\Rightarrow \delta(S^i) = \frac{\beta}{\alpha} (> 0).
\]

If not, for example, if \( \delta(S^i) < \frac{\beta}{\alpha} \), then Axel cannot make up for the lost individual housing unit by producing more composite goods through enhanced pooled production externality net of the congestion externality. The knowledge spillover effect he brings in (less the congestion externality he exerts) is not enough to render the dwindled housing consumption tolerable for the current residents. In this case, city \( i \) is better off bumping him out, i.e., it should reduce \( S^i \), contradicting the fact that \( P \) is in equilibrium. And vice versa, city \( i \) should be larger if \( \delta(S^i) > \frac{\beta}{\alpha} \).

Everyone welcomes Axel and wants more residents to move in in this case. Thus, free mobility arbitrages the gap between externality component \( \delta(S^i) \) and countrywide constant \( \frac{\beta}{\alpha} \) and forces the city to operate in the domain where scale economies are present (otherwise, there will still be an in- or out-flow of people). Note that utility equalization (29) only applies to cities with \( S^i > 0 \).

If a city’s aggregate production function does not exhibit increasing returns to scale anywhere over \( 0 < S^i \leq S \), then the city will not survive and turns rural in the end. In such a location, \( \delta(S^i) < \frac{\beta}{\alpha} \).

---

\( ^{23} \) We took the landlady out of equation because her marginal rate of substitution between composite goods and housing is zero. Cutbacks in housing lot do not affect her at all due to her preferences (38).
and all the residents will be drained off to other cities until $S^i$ becomes zero.\textsuperscript{24} Thus, increasing returns to scale at the aggregate level are an eligibility requirement to be listed under MSA. See appendix A.8 for further discussion on scale economies in $\mathcal{P}$ as opposed to $\mathcal{P}^i$.

It is crucial that we unbundle housing consumption from the composite good. If we include housing as part of a composite good, per capita consumption level, and consequently individual production levels will be the same across the cities because the consumption of leisure is the same everywhere and free mobility guarantees an equal utility level. Then aggregate production level becomes directly proportional to the city size. Alternatively, we can unbundle the composite good and create markets for many commodities. In that case we may have increasing returns to scale in \textit{production} but the price of individual commodities tend to negate the variations in output levels and GMP will be only proportional to the city size. A positive technological shock enhances the production, which reduces the equilibrium price in a perfectly competitive market. Thus, the \textit{value} of the output will exhibit constant returns to scale, which is not compatible with our findings in section 3.2.3. We will have to forgo the assumption of perfectly competitive market in this case.

### 3.3.3 Distribution of GMP

Eekhout [Eeco4] has shown that $S^i$ follows the lognormal distribution using the central limit theorem. The equilibrium size of a city can be written as a sum of the log of error terms over time. The city size depends on the cumulative effect of multiplicative nature rather than of additive nature [LSA01], leading to the lognormal distribution (see appendix A.9 for details). In particular $\log(S^i) \sim N(\mu_S, \sigma_S^2)$. In conjunction with (44) we obtain the following:

\textsuperscript{24} Notice that as the expenditure share of housing $\beta$ increases, it becomes harder and harder to meet the condition (45) and more cities will be abandoned and fewer cities will survive. As we will see later, (46) confirms that a rise in $\beta$ will skew the distribution.
**Proposition 3.3.3: GMP Distribution**

The reported portion of GMP follows a lognormal distribution:

\[
\log Y \sim N \left( \left( \frac{\beta}{\alpha} + 1 \right) \mu_s + \log(\alpha + 2\beta)K, \left( \frac{\beta}{\alpha} + 1 \right)^2 \sigma_s^2 \right)
\]  

(46)

There is a log-linear relationship between GMP and city size (44) and city size follows a lognormal distribution. Naturally, GMP also follows a lognormal distribution by extension. The variance of \( \log(Y) \) is inflated by \( \frac{\beta}{\alpha} + 1 \) due to citywide scale economies (proposition 3.3.2). This observation is consistent with our findings in section 3.2.2. GMP (in log scale) spreads further than its city-size counterpart in \( P \).

Eeckhout [Eeco4] also establishes the Gibrat’s law \( \frac{d \log S_i}{dt} \approx \epsilon \) (\( t \) denotes time. See appendix A.9). Then from (44),

\[
\frac{d \log Y_i}{dt} = \left( \frac{\beta}{\alpha} + 1 \right) \frac{d \log S_i}{dt} \approx \left( \frac{\beta}{\alpha} + 1 \right) \epsilon_i^{\prime},
\]

where \( \epsilon_i^{\prime} \) is an i.i.d. random variable. Thus GMP also follows the Gibrat’s law. Note that variation in \( Y^i \) is inflated by \( \frac{\beta}{\alpha} + 1 \) and this is coherent with our empirical findings in section 3.4.1.

### 3.4 Empirical Implementation

#### 3.4.1 Distribution of GMP

We will put proposition 3.3.3 to an empirical test in this section. First, rewrite the GMP distribution (46) as \( \log Y \sim N(\mu, \sigma^2) \). The maximum likelihood estimator of (46) is \( \hat{\mu} = \frac{\sum \log Y_i}{N} \) and \( \hat{\sigma}^2 = \frac{\sum (\log Y_i - \hat{\mu})^2}{N} \). We report our estimations in table 11 with supporting density plots in figure 27. Housing portion of GMP, \( \beta K \left( S^i \right)^{\frac{\beta}{\alpha} + 1} \), is also available and they are expected to follow the lognormal distribution as well.
(a) PDF of GMP

(b) PDF of Housing

(c) CDF of GMP

(d) CDF of Housing

(e) PP Plot of GMP

(f) PP Plot of Housing

Figure 27.
Employment & GMP & Housing

<table>
<thead>
<tr>
<th></th>
<th>Employment</th>
<th>GMP</th>
<th>Housing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Size I</td>
<td>366</td>
<td>366</td>
<td>323</td>
</tr>
<tr>
<td>Censored (Unreported Cities)</td>
<td>0%</td>
<td>0%</td>
<td>11.75%</td>
</tr>
<tr>
<td>Censored (Unreported Value)</td>
<td>0%</td>
<td>0%</td>
<td>13.87%</td>
</tr>
<tr>
<td>Estimated Mean ( \hat{\mu}_S, \hat{\mu} )</td>
<td>12.68</td>
<td>23.13</td>
<td>20.66</td>
</tr>
<tr>
<td>Estimated Variance ( \hat{\sigma}_S^2, \hat{\sigma}^2 )</td>
<td>1.128</td>
<td>1.478</td>
<td>2.133</td>
</tr>
<tr>
<td>Theoretical Variance ( \left( \frac{\hat{\beta}}{\hat{\alpha}} + 1 \right) \hat{\sigma}_S^2 )</td>
<td>–</td>
<td>1.571</td>
<td>1.571</td>
</tr>
<tr>
<td>Log Likelihood/( I )</td>
<td>.2709</td>
<td>.2370</td>
<td>.1968</td>
</tr>
<tr>
<td>Kolomogorov-Smirnov Statistic</td>
<td>.1122</td>
<td>.1032</td>
<td>.1031</td>
</tr>
</tbody>
</table>

Table 11.

The overall fit is not too far off. The maximum discrepancy between the empirical and estimated CDF (Kolomogorov-Smirnov statistic) is .1032. We are more than certain that the fit would improve if we used an inclusive data set. Due to truncation to the left of the distribution, the tail end of the distribution does not extend as far as the theory predicts to the left. At this point we do not have GMP data for smaller cities. Hopefully GMP on micropolitan statistical areas or census-designated places will become available someday but we will settle for MSA data for now.

As we will see in section 3.4.2, the ratio between \( \beta \) and \( \alpha \) is .1800. Following the GMP distribution (46), the theoretically expected value of \( \sigma^2 \) is \( (1.180\hat{\sigma}_S^2)^2 = 1.571 \) (cf. theoretical variance in table 11). Our expected variance in GMP computed from the expected expenditure shares \( \alpha \) and \( \beta \) and estimated variance \( \sigma^2 \) in employment is eerily close to the actual variance in GMP (we missed the actual value only by 5.86%). This confirms the validity of the form (46) along with the selection of the utility and production functions in section 3.3.2. On the other hand, the variance on housing is larger than the theoretical value by 35.86%. We will explore the cause of the large gap in housing in section 3.4.2.

3.4.2 Scale Economies

According to (44), the ratio between GMP and housing is \( \alpha+2\beta \) to \( \beta \). The actual ratio is $1.162e+13$ to $1.603e+12$ among the MSA’s, indicating that the expenditure share \( \beta \) of the housing sector is
13.23%.\(^{25}\) Hence, the expected ratio \(\frac{\beta}{\alpha} = 0.1800\). Taking a log of (44),

\[
\log Y_i = \log(\alpha + 2\beta)K + \left(\frac{\beta}{\alpha} + 1\right) \log(S_i) = \log(\alpha + 2\beta)K + 1.1800 \log(S_i).
\]

The actual value of the coefficient is 1.117 in Table 10 rather than 1.180, meaning that our model overshoots the coefficient by only 5.64% (or, the economy is off where it should be by 5.64%). The housing portion of GMP \(\log\beta K(\frac{S_i}{\alpha} + 1)\) also shares the same coefficient in (44). Here the predicted value comes short of the actual value by 10.59%. The large discrepancy may be because of the censored data,\(^{26}\) or the landlords’ or developers’ registered addresses, which may be different from the city where they have their real estate. The fact that imputed rent is excluded and that houses usually last longer than the duration of a fiscal year exacerbates the deviation even further.

### 3.5 Conclusion and Extensions

We have discovered that the GMP distribution shares the same pattern as the city-size distribution, and we sought a systematic illustration of how our local economies are related to their employment and output levels on a national scale. Proposition 3.3.2 further revealed that GMP grows more than 1% against 1% growth in employment. Large cities make up for an exceeding share of GDP and they do so more than their city size alone can account for. Consequently, due to agglomeration economies of GMP in employment, the GMP distribution is even more skewed than the city-size distribution.

We constructed a production economy model that endogenously gives rise to agglomeration economies in equilibrium. The interplay between externalities and housing consumption drives the cities to operate at the size where increasing returns to scale are present. Empirical testing

\(^{25}\)Note that GDP includes real estate sales and excludes imputed rent. Thus, the figure does not necessarily represent the expenditure share in a particular year.

\(^{26}\)Housing output values are not reported for all MSA’s and the missed portion is not negligible. For example, housing sector in Dallas (rank #4 in employment) is censored out of the data. However, these censored values are included in the nationwide aggregate.
verifies our model prediction; however, the result could be made better off. Due to data truncation, our predicted distribution does not trace the lower end of the distribution well. Ideally, we would like to test our prediction with an exhaustive data set, which, for the moment, does not exist yet.

Our objective was to explain the GMP distribution in a consistent manner. As far as we know this is the first attempt to analyze GMP as a distribution. Along the way, we have left several prospects for future work. We assumed a single-input production function. Local output may well be affected by capital, educational attainment, location of the city, access to a rich labor pool, or urban infrastructure, which, obviously vary from city to city. We also packed the consumption goods other than housing and leisure into a single basket. In reality a city comes with various industries. Some of them may exhibit increasing returns to scale and some may not both within and across the industries. Cross-sectional GMP analysis is called for to decode the internal workings of local economies that, on aggregate, exhibits increasing returns to scale and a fat-tail distribution. Lastly, we assumed that all the goods are immobile. Data reported by the Bureau of Economic Analysis are based on tax filed in each MSA. As such, the scope of GMP matches the range of production in MSA, but it does not necessarily match the range of consumption in the city. The openness of a spatial economy may be addressed by adding shipping firms to the production economy. One way to do so is to assume that a shipping firm takes composite goods in city $i$ as input and “produces” the same composite good in city $j$ as output, in less than a one-to-one ratio to reflect shipping charges of iceberg form.
A  APPENDIX

A.1  A Model from the Literature Modified to Include Insurance

A.1.1  Notation

We use the model of Eeckhout [Eec04] as the basis for the analysis because it is explicit about consumer behavior, in the form of an optimization problem, as well as endogenous urban variables, namely local wages and land rents.

The original model is specified as follows. For complete detail, see Eeckhout ([Eec04], pp. 1445-1446). In general, there is a large number of cities and a continuum of identical consumers. Each city produces the same commodity using labor and a constant returns to scale technology. The production function is dependent on a city-wide shock and on a positive agglomeration externality that is a function of city population. There is also a negative congestion externality that is a function of city population and that only affects consumers. On net, the random shocks to productivity cause some, but not all, population to move each period so as to equalize utility across cities in equilibrium.

Time is discrete and indexed by $t$. The set of cities is indexed by $i \in I$. Consumers are infinitely lived and identical. In city $i$ at time $t$, consumption good is $c_{i,t}$, housing or land consumption is $h_{i,t}$ whereas leisure is $1 - l_{i,t}$ for labor supply $l_{i,t} \in [0,1]$. Utility for a consumer in city $i$ at time $t$ is Cobb-Douglas:

$$u(c_{i,t}, h_{i,t}, l_{i,t}) = c_i^\alpha h_i^\beta (1 - l_i)^{1-\alpha-\beta}$$

with $\alpha, \beta, \alpha + \beta \in (0,1)$.

Production is constant returns to scale. The measure of population in city $i$ at time $t$ is $S_{i,t}$. Let $A_{i,t}$ be the technological productivity parameter of city $i$ at time $t$. This parameter follows the law of motion:

$$A_{i,t} = A_{i,t-1}(1 + \sigma_{i,t})$$ (47)
where $\sigma_{i,t}$ is the exogenous technological shock to city $i$ at time $t$. It is assumed that $\sigma_{i,t}$ is i.i.d. with mean 0, symmetrically distributed, and satisfies $1 + \sigma_{i,t} > 0$. The positive local externality (spillover) function is given by $a_+(S_{i,t}) > 0$, where $a'_+(S_{i,t}) > 0$. The marginal product of a worker in city $i$ at time $t$ is given by

$$y_{i,t} = A_{i,t}a_+(S_{i,t})$$

For prices, let the consumption good be numéraire, the price of housing or land in city $i$ at time $t$ be $p_{i,t}$, and let the wage in city $i$ at time $t$ be $w_{i,t}$. The local negative externality or congestion function is given by $a_-(S_{i,t}) \in [0,1]$, where $a'_-(S_{i,t}) < 0$. The optimization problem of a consumer in city $i$ at time $t$ is:

$$\max \left\{ c_{i,t}, h_{i,t}, l_{i,t} \right\}$$

subject to

$$c_{i,t} + p_{i,t}h_{i,t} \leq w_{i,t}L_{i,t}$$

where $w_{i,t} = A_{i,t}a_+(S_{i,t})$ and $L_{i,t} = a_-(S_{i,t})l_{i,t}$. Total land or housing in a city is $H$.

Using the first order conditions from this optimization problem and market clearance, equilibrium (denoted by asterisks) in city $i$ at time $t$ as a function of population $S_{i,t}$ can be found:

$$p^*_{i,t} = \frac{\beta A_{i,t}a_+(S_{i,t})a_-(S_{i,t})S_{i,t}}{H}$$

$$w^*_{i,t} = A_{i,t}a_+(S_{i,t})$$

$$c^*_{i,t} = \alpha A_{i,t}a_+(S_{i,t})a_-(S_{i,t})$$

$$h^*_{i,t} = \frac{H}{S_{i,t}}$$

$$l^*_{i,t} = \alpha + \beta$$

The last equation in particular, indicating that labor supply is independent of population, is an artifact of the Cobb-Douglas specification.
Substituting back into the utility function, indirect equilibrium utility as a function of population $u^*(S_{i,t})$ can be written as

$$u^*(S_{i,t}) = [\alpha A_{i,t} \cdot a_+(S_{i,t})a_-(S_{i,t})]^\alpha S_{i,t}^{-\beta} H_{i,t}^{\beta} [1 - \alpha - \beta]^{1-\alpha-\beta}$$  \hspace{1cm} (48)$$

Under free mobility of consumers, indirect utility is equated across cities in each time period, determining their populations as a function of their productivity and their realized history of shocks, summarized by $A_{i,t}$. Instantaneous utility is constant over both time and location in equilibrium. Again using Eeckhout’s notation, call this instantaneous utility level $U$.

A.1.2 Insurance

Let the discount factor be denoted by $\rho \in (0,1]$. In correspondence with the assumption of complete capital markets, it is assumed that all consumers can borrow or lend at rate $\frac{1}{\rho} - 1$. The consumer optimization problem (at time $0$) becomes:

$$\max_{\{c_{i,t}, h_{i,t}, l_{i,t}\}} \sum_{t=1}^{\infty} \rho^t \cdot c_{i,t}^{\alpha} h_{i,t}^{\beta} (1 - l_{i,t})^{1-\alpha-\beta} \hspace{1cm} \text{subject to} \hspace{1cm} \sum_{t=1}^{\infty} \rho^t \cdot (c_{i,t} + p_{i,t} h_{i,t}) \leq \sum_{t=1}^{\infty} \rho^t \cdot w_{i,t} L_{i,t}$$

As stated by Eeckhout, the problem reduces to the one period optimization problem if there are no insurance or futures markets. Formally, there should be an expectation in the objective function and a requirement that the budget constraint hold for every state of nature. However, this is omitted in the literature since the problem is reduced to a static optimization problem where the state of nature is observed before consumers make their choices.

There are several important points to be made at this juncture. First, it is useful to imagine the consumers stepping back at $t = 0$ and making decisions about their cities of residence and
their consumption bundles for the entire time stream of their infinite lives, contingent on state realizations at each time. Second, and more important, it does not matter which interpretation of the model one employs. Specifically, resources can be transferred across states of the world (at any given time) in one or more of several ways (insurance, self-insurance, or futures contracts). In the end, what a consumer is choosing is their residence and consumption bundle for every time and for every possible state of the world, optimizing utility subject to the budget constraint. The state of the world at time \( t \) affects the optimization problem through the prices, \( p_{i,t} \) and \( w_{i,t} \), and income (through \( a_-(S_{i,t}) \) and \( L_{i,t} \)) only. These variables depend on \( A_{i,t} \) both directly and indirectly, the latter because \( S_{i,t} \) depends on \( A_{i,t} \) in equilibrium. The state of the world at time \( t \) does not enter into the consumer optimization problem otherwise. For example, it does not enter into the utility function. We could index these prices and incomes by the state of the world, but that would only serve to complicate notation.

As already mentioned, what will matter are only the lifetime choices of residence and consumption bundles, contingent on the state of the world in each period. The method used to actually implement them, via transfers across states in a time period as opposed to across time periods, does not matter; there are many possibilities. With complete futures markets, at time \( t = 0 \) the consumers can sell their labor in every future time period and state, buying consumption good and housing in every future time period and state. With insurance markets, at \( t = 0 \) the consumers can buy actuarially fair insurance against price and income changes. With self-insurance, they can commit to a plan of borrowing and saving under all possible scenarios, namely realizations of states in each time period.

To get the basic idea across, in the next subsection we show how insurance would work from the beginning when all cities have the same initial state (productivity) and population. This yields no movement at any time in equilibrium. In the next subsection, we discuss how to extend this so that insurance can begin from equilibrium of the model at any time \( t \). From that time on, there is no consumer movement unless the insurance is switched off.
Insurance when the initial state is the same for all cities

To illustrate the ideas behind insurance, we begin with an example where all cities start with the same state at time 0 and consumers insure from then on.

For notational purposes, let $\bar{S}$ be the mean population of cities, that is $\bar{S} = \frac{\sum_{i \in I} S_i}{|I|}$, where $|I|$ is the cardinality of the set $I$. Let $A_0 = A_{i,0}$ denote the common initial technology level for all the identical cities before the process begins. Let $S_{i,0} = \bar{S}$ for all cities $i$, so they all have the same initial population. We assume that

$$U = u^*(\bar{S}) = [\alpha A_0 \cdot a_+(\bar{S})a_- (\bar{S})]^\alpha \bar{S}^{-\beta} H^\beta [1 - \alpha - \beta]^{1-\alpha-\beta}$$

Thus, we assume for illustrative purposes that the initial configuration of shock $A_0$ and uniform population distribution $\bar{S}$ generate the instantaneous equilibrium utility. This is to get the idea across; in the next section, we will show how to start insurance from equilibrium at an arbitrary given time. In either case, no consumer movement will occur once insurance begins.

With insurance, self-insurance, or a futures market (or some combination of all 3), we propose the following equilibrium solution for all cities $i$ and times $t$: 

$$\bar{p}_{i,t} = \frac{\beta A_0 a_+(\bar{S})a_- (\bar{S})\bar{S}}{H}$$

$$\bar{w}_{i,t} = A_0 a_+(\bar{S})$$

$$\bar{c}_{i,t} = \alpha A_0 a_+(\bar{S})a_- (\bar{S})$$

$$\bar{h}_{i,t} = \frac{H}{\bar{S}}$$

$$\bar{I}_{i,t} = \alpha + \beta$$

In other words, this is the allocation generated by a constant, over both time and state, allocation with a uniform distribution of consumers. By construction, it generates the same instantaneous utility stream for all consumers in all cities and in all times as both the initial distribution and the equilibrium studied by Eeckhout.
But how does this work in a pragmatic sense? Regarding futures markets, each consumer works the same hours, independent of state. If the state realization is good, i.e. if the consumer is in city $i$ at time $0$ and $A_{i,t} > A_0$, income in excess of $A_0a_+(\bar{S}) \cdot (\alpha + \beta)$ is paid to the market. If the state realization is bad, then the consumer receives income from the market, smoothing consumption. Under self-insurance, the consumer commits to a plan of saving income in a good state, and withdrawing from savings or borrowing in a bad state, thus smoothing consumption. The banks know that $E(A_{i,t}) = A_0$, so they are willing to lend. Under mutual insurance, the same type of idea, with commitment, has consumers who are in cities with good states at time $t$ contributing to an insurance pool, and those in cities with bad states receiving payments from an insurance pool. If the number of cities is large, the law of large numbers implies that the mutual insurance pool is solvent.

It is interesting to note that the phenomenon we describe is something like another manifestation of Starrett’s spatial impossibility theorem (see Mills [Mil67], Starrett [Sta78], Fujita [Fuj86], and Fujita and Thisse [FT02] (chapter 2.3), though here markets are incomplete due to the presence of unpriced local externalities, both positive ($a_+$) and negative ($a_-$). In particular, we obtain a uniform distribution of economic activity, in spite of the violation of one of the hypotheses of the Theorem, namely perfect and complete markets. It is well-known (from these cites) that the hypotheses of Starrett’s Theorem are sufficient but not necessary for the conclusion, namely the lack of agglomeration.

In summary, the equilibrium time path of utility for every consumer is the same, and constant, under insurance and under the equilibrium that generates movement and eventually becomes lognormal. At the very least, a discussion of why the latter equilibrium is selected should be offered in the literature.

With any moving cost, the insurance or futures market equilibrium (the one denoted with bars) clearly dominates the path with asterisks, the one put forth in the literature. Given a choice between moving along the equilibrium path or insuring at $t = 0$, each consumer will individually choose to insure.
A second, and perhaps more reasonable possibility, is that consumers observe $A_{i,t}$ imperfectly when they make their location decisions each period. In that case as well, the consumers will insure rather than move, since they are risk averse. This can be seen in equation (48). When consumers cannot perfectly observe $A_{i,t}$ when optimizing, equilibrium expected utility will vary in proportion to $E(A_{i,t})^a$.

**Insurance starting when the state is an equilibrium at a given time**

The preceding subsection was provided to give intuition. However, it has drawbacks in terms of commitment on the part of consumers if they use mutual insurance at each given time, and on the part of banks and consumers at time 0 if the consumers use self-insurance. Moreover, there is a strong assumption that at time 0, $A_0$ is the same across cities, each city has the same population $S$, and this combination produces the instantaneous equilibrium utility level. Here we discuss how to dispense with some of these assumptions.

Suppose that we start running the model without insurance, so that consumers are generally moving around, and stop it at some arbitrary time $t$. At this time, the instantaneous utility level of each consumer is, of course, $U$. Consider a consumer in city $i$ and the possibility of self-insurance. At that point, the productivity parameter in the city is $A_{i,t}$, and everyone knows from equation (47) that for $t' > t$, $E(A_{i,t'}) = A_{i,t}$. So if the consumers in that city freeze their consumption bundle at whatever it is at that time, and commit to staying in that city and consuming that consumption bundle forever through a plan of borrowing and saving, they will obtain utility level $U$ in each period. This exploits the law of large numbers over time.

Mutual insurance, exploiting the law of large numbers over space at a given time, is more interesting. Pick an arbitrary time $t$ and freeze all the consumers in their equilibrium locations as well as their consumption bundles. All consumers obtain utility $U$ in this situation at time $t$. Now consider what would happen if they maintain the same location and consumption bundle in time $t + 1$. Given equation (47), the surplus or deficit in total wage payments for city $i$ relative
to the benchmark inherited from the previous period \( t \) is

\[
\sigma_{i,t+1} \cdot A_{i,t} \cdot a_+(S_{i,t}) \cdot (\alpha + \beta) \cdot S_{i,t}
\]

(49)

Thus, to ensure that this system of mutual insurance across cities is solvent at time \( t + 1 \), it is necessary that

\[
\sum_{i \in I} \frac{1}{|I|} \cdot \sigma_{i,t+1} \cdot A_{i,t} \cdot a_+(S_{i,t}) \cdot (\alpha + \beta) \cdot S_{i,t} = 0
\]

Although this cannot be assured for finite \( |I| \), we can see that as the number of cities \( |I| \) tends to infinity, the limiting result is a consequence of a law of large numbers with weights given by

\[
\frac{1}{|I|} \cdot A_{i,t} \cdot a_+(S_{i,t}) \cdot (\alpha + \beta) \cdot S_{i,t}.
\]

Since the support of the random variable \( 1 + \sigma_{i,t} \) is contained in \((0,2)\), equation (47) implies that the size of \( A_{i,t} \) at given time \( t \) can be bounded over \( i \) by \( 2^t A_{i,0} \). Since \( A_{i,t} \) and \( S_{i,t} \) are positively related, there is also a bound for \( S_{i,t} \) and thus for the continuous function \( a_+(S_{i,t}) \) for fixed \( t \) over \( i \).

There is an extensive literature on law of large numbers for sums of weighted random variables. Our framework would fit, for example, in Cabrera and Volodin (2005, Corollary 1).

Notice that there is no commitment required under mutual insurance beyond the next period. So it can be switched on and off as desired, with no consumer movement when it is on, and movement when it is off. If insurance is carried on to period \( t + 2 \), then expression (49) updated to time \( t + 2 \) represents the change in the surplus or deficit in total wage payments for city \( i \) relative to time \( t + 1 \), so solvency at time \( t + 2 \) requires that these changes sum to zero across cities.

There is a subtle issue of commitment relevant to mutual insurance that is not as subtle for self-insurance. An interesting strategy for consumers is to insure for the first period, wait for uncertainty in the first period to be resolved, then pick a winner (a location where \( A_{i,t+1} > A_{i,t} \)), move there, and then reinsure. If a long term commitment to insurance is required, then of course (similar to self-insurance) this strategy is not possible. Alternatively, if such commitment is not desirable, we could make an assumption about off the equilibrium path beliefs. That is,
we could assume that if a consumer moves to a winning location, that person thinks that others will also follow this strategy, driving down instantaneous utility until it is again equalized across locations. Thus, this strategy does not do better than insuring in every period.

**Extensions: Partial Insurance and Moving Cost**

Before discussing partial insurance, it is useful at this point to make some remarks and to give some detail about equilibrium in Eeckhout’s model with moving costs. In such a model, equilibria do not feature equalization of utility across locations, and this makes matters more complicated. We assume that a moving cost, if non-zero, is paid in terms of the numéraire consumption good. We call it $M \geq 0$. The main fact we need is that the explicit utility function is strongly monotonic in consumption good.

In this framework, we assume that before uncertainty is realized in a given period, each consumer must choose between a commitment to stay in their current location (and possibly insure against the uncertainty) or to move (after the realizations of all shocks that period are known to everyone).

First, consider the situation where the initial distribution generates the same instantaneous utility, independent of location. (With moving cost, this might not happen, but it is useful for thinking about the problem.) Then we will consider the situation where the initial distribution does not generate the same instantaneous utility.

The first claim we make is that under these conditions, for fixed positive $M$, every consumer will unilaterally decide to commit to stay and insure. The reason is that for Eeckhout’s arguments to work, uncertainty must be arbitrarily small. This is made explicit p. 1447 of Eeckhout (2004), where $\varepsilon_{i,t}$ is written as an increasing function of $\sigma_{i,t}$, and asymptotic statements as $\varepsilon_{i,t} \to 0$ are used to derive Gibrat’s law. So for any fixed $M > 0$, we can find sufficiently small $\varepsilon_{i,t}$, and thus $\sigma_{i,t}$ (or its support) so that each consumer would find it more costly to move than to insure. Thus, the potential gain from moving will be less than the cost, namely $M$. If consumers have heterogeneous moving costs, as long as there isn’t an atom of consumers at zero moving cost,
as the random variables representing the shocks become small, the measure of consumers who prefer to move rather than insure will tend to zero.

Clearly, in the original equilibrium with consumer movement, the measure of consumers who actually move between periods to equate utility across locations is relatively small. But given the last argument, no individual will want to be among the movers. And over time, the social cost of moving will add up.

Given that nobody chooses to move, we can next calculate the amount of insurance they will purchase. Since $S_{i,t+1} = S_{i,t}$, we can calculate from (48) and (47):

$$Eu^*(A_{i,t+1} | A_{i,t}, S_{i,t}) = E(\alpha A_{i,t}(1 + \delta \cdot \sigma_{i,t+1})) \cdot a_+(S_{i,t})a_-(S_{i,t})^{\beta} [S_{i,t}]^{-\beta} H^{\beta} [1 - \alpha - \beta]^{-\alpha - \beta}$$

where $1 - \delta$ is the percentage of insurance purchased. By concavity, $E((1 + \delta \cdot \sigma_{i,t+1})^\alpha)$ is maximized at $\delta = 0$, so everyone purchases full insurance. Partial insurance is not a feature of equilibrium.

Finally, consider the case where the initial indirect utility levels are not equalized across locations in equilibrium due to the moving cost. In fact, due to costly mobility, the utility difference in terms of numéraire cannot exceed $M$, for otherwise consumers would move to the location with higher gross utility (as it exactly compensates for the moving cost). But since consumers can look ahead, they realize that this is not just a one period difference in utility levels. In other words, after sinking moving cost this period, next period it is expected that the higher utility location will yield the same utility (for example, by fully insuring) without having to pay the moving cost. So more people will move there this period. The present discounted value of the utility difference will be at most the moving cost. So in the first period under consideration, people might move. But after that, the population distribution will be stable, and everyone in all locations will fully insure for the same reason as given above.
A.2 Positive Transport Costs

Before turning to the details of our simulations, we first summarize our conclusions. Table 12 reports the statistics from our simulations. Typically, in equilibrium, there will be more than one city producing the same commodity to serve nearby cities. Whereas both the lognormal and GEV distributions track the simulated data well, the results are inconclusive concerning which of the two distributions better explains the economy with multiple production sites. The KS statistic is small for both distributions; AIC and BIC are lower for the lognormal distribution when the iceberg transportation cost is large, but in the end the differences are small.
<table>
<thead>
<tr>
<th>Iceberg</th>
<th># of Cities</th>
<th>Distribution</th>
<th>$\hat{\mu}$</th>
<th>$\hat{\sigma}$</th>
<th>$\hat{\xi}$</th>
<th>$\hat{\kappa}$</th>
<th>$\hat{\gamma}$</th>
<th>Log likelihood</th>
<th>AIC</th>
<th>BIC</th>
<th>KS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.01</td>
<td>78/300</td>
<td>Lognormal</td>
<td>-4.624</td>
<td>0.8466</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>-275.2</td>
<td>554.4</td>
<td>559.1</td>
<td>8.343E-02</td>
</tr>
<tr>
<td>1.01</td>
<td>78/300</td>
<td>GEV</td>
<td>4.845E-02</td>
<td>1.051E-02</td>
<td>-0.4621</td>
<td>1233</td>
<td>0.7967</td>
<td>-267.8</td>
<td>545.5</td>
<td>557.3</td>
<td>7.152E-02</td>
</tr>
<tr>
<td>1.02</td>
<td>107/300</td>
<td>Lognormal</td>
<td>-5.250</td>
<td>1.1704</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>-345.2</td>
<td>694.5</td>
<td>699.8</td>
<td>4.886E-02</td>
</tr>
<tr>
<td>1.02</td>
<td>107/300</td>
<td>GEV</td>
<td>4.240E-02</td>
<td>1.711E-02</td>
<td>-0.04938</td>
<td>27.60</td>
<td>0.7377</td>
<td>-346.1</td>
<td>702.2</td>
<td>715.5</td>
<td>5.477E-02</td>
</tr>
<tr>
<td>1.05</td>
<td>92/300</td>
<td>Lognormal</td>
<td>-5.301</td>
<td>1.3861</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>-307.5</td>
<td>618.9</td>
<td>623.9</td>
<td>1.062E-01</td>
</tr>
<tr>
<td>1.05</td>
<td>92/300</td>
<td>GEV</td>
<td>4.589E-02</td>
<td>1.491E-02</td>
<td>-0.1759</td>
<td>1839</td>
<td>0.8107</td>
<td>-310.7</td>
<td>631.4</td>
<td>644.0</td>
<td>1.028E-01</td>
</tr>
<tr>
<td>1.1</td>
<td>93/300</td>
<td>Lognormal</td>
<td>-5.345</td>
<td>1.4345</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>-319.1</td>
<td>642.2</td>
<td>647.4</td>
<td>8.516E-02</td>
</tr>
<tr>
<td>1.1</td>
<td>93/300</td>
<td>GEV</td>
<td>1.456E-01</td>
<td>3.837E-02</td>
<td>-0.3230</td>
<td>250.5</td>
<td>0.8530</td>
<td>-321.6</td>
<td>653.2</td>
<td>666.0</td>
<td>8.148E-02</td>
</tr>
</tbody>
</table>

Table 12.
Now let us turn to the details of the simulations. We have 10 industries with 30 potential production sites each, so there are 300 potential city locations lined up on a circle. Transportation cost is of the iceberg form. In particular, we ship out 1.01, 1.05, or 1.1 units of commodity and an immediate neighbor on the circle receives 1 unit of it. Shipment cost grows with distance traveled. City residents purchase goods from the city that quotes the smallest delivered price within each industry. Everything else is the same as the basic model we have described.

The results from our simulations depart from the results from the basic model in two ways. As is the case for the basic model, high productivity reduces the mill price. However, it is not practical to serve the entire population from a single city just because the city is the most productive in the industry; delivered price grows with distance. Rather, a couple of cities will coexist within the same industry to serve the cities in close proximity to each of them. Also, surviving cities are not necessarily the most productive in the industry. It depends not only on a city’s own productivity but also on the size and productivity of close neighbors.

For the 300 locations, we mark an industry by \( j = 1, \ldots, 10 \) and a location block of 10 cities by \( i = 1, \ldots, 30 \). On a circle, we have city \((i, j) = (1,1)\) located next to \((1,2)\), next to \((1,3)\) and so on. Then, next to \((1,10)\) we have \((2,1)\). As we travel along the circle in this way, we will eventually reach \((30,10)\), whose next neighbor is \((1,1)\), completing the circle. Figure 28 is a schematic representation of the arrangement of cities so that we can illustrate how the model works. A circle with 300 dots would be difficult to interpret, so instead we give the industry on the horizontal axis and the location block on the vertical axis.

Figure 28(a) represents the size of the random productivity draw \( A \), where a larger dot represents a larger value of \( A \). We shall use this single draw of the random variable for subsequent illustrations. In figure 28(b) we represent the equilibrium city-size distribution for a transport cost of 1.01, where a larger dot means a larger equilibrium population and no dot means the area is rural. Such a small transport cost causes only a minor change from our basic model. Take industry \( j = 8 \) for example. The most productive city is \((i, j) = (6,8)\). This would be the city that survives in the absence of any transport cost. City \((6,8)\) still survives and produces most
Figure 28. Missing dots indicate rural areas in figures 28(b) and 28(c).

of commodity $j = 8$ when we throw in a minimal shipment fee. However, there are some other cities, such as $(25, 8)$, that also engage in production to serve local markets. City $(25, 8)$ undercuts the delivered price of city $(6, 8)$ for nearby cities. Indeed, shipment from $(6, 8)$ to $(25, 8)$ would require impractical 189 steps.

As we raise transport cost, a qualitatively different city-size distribution emerges in equilibrium. Own productivity level becomes less pertinent and the size of neighboring cities becomes
more influential in determining city size. This can be seen in figure 28(c), where transport cost is raised to 1.1 units. Take industry 8 again. The most productive city (6,8) is still in the picture but production is more intense in the less productive city (17,8) when transport cost is raised. City (6,8) is quite productive in isolation, but is surrounded by cities whose productivity is not exceptional. Consequently, its local market is small. On the other hand, city (17,8) is surrounded by productive cities. Since utility is concave, there is large demand for commodity 8 from these productive (and therefore, populous) neighbors. And this large demand will be fulfilled by the nearby city (17,8) rather than a faraway city (6,8) to ward off the increased shipping charge. City (17,8) grows to support local demand that, in turn, will create a large demand for goods other than 8 produced by its neighbors. As a result, high transportation cost creates snowballing clusters of cities, whose average productivity across the industries within the region is high, and eliminates cities of high productivity in geographic isolation. The equilibrium does not simply select the most productive cities as survivors.

The various theories we have summarized have zero transport costs. The simulations indicate that positive transport costs can generate a new force in city selection, namely a kind of local market effect illustrated in figure 28(c).

Figure 29 and table 12 summarize maximum likelihood estimation for the lognormal and GEV city size distributions from these simulations. On the whole, both distributions fit the simulations well. Notice that the number of active cities is not monotonically increasing in transport cost. The reason is that some clusters of cities empty out as transport cost increases.

Table 12 reports the estimated parameters. The last four columns are the values of the log likelihood function (larger is better), Akaike and Bayesian Information Criteria (AIC and BIC), and the Kolmogorov-Smirnov (KS) statistic (smaller is better for all). Judging by the values of the KS statistic, the overall fit is not bad. Turning next to the comparison between the lognormal and GEV distributions, the simulated city size distributions yield better log likelihood values for GEV when transport cost is small but favor lognormal for AIC and BIC (GEV has more parameters). In the end, the fits are almost identical. In additional simulations not detailed here, we found
no systematic relationship between transportation cost and how well either distribution fits the data.

Figure 29. Maximum likelihood estimation with iceberg transportation parameter (from top row) 1.01, 1.02, 1.05 and 1.1. Transportation becomes costlier in this order.
A.3 Proof of Proposition 2.3.1

Proof. Note that $s(a_i)$ is monotone decreasing in $a_i$. Suppose $J > 2$ and the network is neither complete or completely isolated. We have

$$s'(a_i) := \frac{ds(a_i)}{da_i} = -(\log \tau)s(a_i)S^{-1}(S - s_i) \leq 0$$

with equality iff $\tau = 1$. The second derivative is, therefore,

$$\frac{d^2s(a_i)}{da_i^2} = [s'(a_i)]^2 \frac{S - 2s_i}{s_i(S - s_i)} \geq 0,$$

with equality iff $\tau = 1$. Hence $s(a_i)$ is strictly convex in $a_i$.

To show that $s(a_i)$ bulges as $\tau$ grows, first note $\frac{\partial s(a_i)}{\partial \tau} = -\tau^{-1}s(a_i)(a_i - AB^{-1})$, where $A := \sum_j a_j \tau^{-a_j}$ and $B := \sum_j \tau^{-a_j}$. Then

$$\frac{dD(\tau)}{d\tau} = \frac{1}{2\tau} \left\{ [s(a_M) - s(a_H)](a_H - AB^{-1}) + [s(a_M) - s(a_L)](a_L - AB^{-1}) \right\},$$

where $a_M := (a_H + a_L)/2$. The first term in the curly braces is positive because $s(a_M) - s(a_H) > 0$ and $a_H - AB^{-1} = B^{-1}\sum_{j \neq H}(a_H - a_j)\tau^{-a_j} > 0$. Likewise, the second term is positive because $s(a_M) - s(a_L) < 0$ and $a_L - AB^{-1} < 0$. Therefore $\frac{dD(\tau)}{d\tau} > 0$, which establishes the claim. $\Box$

A.4 Idea behind Geodesic Length (16)

We briefly repeat [HSF$^+$05]'s arguments to obtain (16) in our context. Consider a geodesic between nodes $v_i$ and $v_j$. We ignore loops. The probability that a child node traces back to its ancestors via some circumvention is proportional to $1/J$. It becomes negligible as the system size $J$ grows (our system size ranges from 69 to 25,358 in section 2.4). As shown in [HSF$^+$05], the resulting error is minimal. A tree is a sequence of nodes where each node except for the root node has exactly one parent (or ancestor) node. Each node may or may not be followed by (a) child
node(s). There are no cycles on a tree. If we pick a random tree starting from \( v_i \), we will wind up at \( v_j \) somewhere along the tree \( k_j/\sum_{r\in V} k_r \) of the time and we will not reach \( v_i \) the remaining \( 1 - k_j/\sum_{r} k_r \) of the time. On average, we will reach \( v_j \) within \( \sum_r k_r/k_j \) trials. Suppose that the depth (the number of parent nodes that you have to go through before reaching your root node) of \( v_j \) is \( l \). There are \( k_i \kappa^{l-1} \) nodes whose depth is \( l \). Therefore, on average, we arrive at \( v_j \) in \( l \) steps if

\[
\frac{\sum_r k_r}{k_j} = k_i \kappa^{l-1},
\]

from which we obtain (16). In other words, if, on average, it takes more than \( k_i \kappa^{l-1} \) trials to reach city \( j \), i.e., \( \frac{\sum_r k_r}{k_j} > k_i \kappa^{l-1} \), then it is likely that city \( j \) is more than \( l \) steps away from your city \( i \). You would try \( k_i \kappa^{l-1} \) times to find city \( j \), when in fact you would need additional \( \frac{\sum_r k_r}{k_j} - k_i \kappa^{l-1} \) trials to reach city \( j \), meaning that city \( j \) is not in the group of cities \( l \) steps away from you but actually located somewhere farther down. On the contrary if it takes less than \( k_i \kappa^{l-1} \) trials to reach city \( j \), then city \( j \) should be less than \( l \) steps away from you. You would not need that many trials to find a city \( j \), the implication being that, once again, you are looking at a wrong group of cities. Thus, city \( i \) and \( j \) are \( l \) steps apart from each other exactly when (50) is satisfied with equality.

### A.5 Branching Factor

Take a random edge and walk towards one arbitrarily selected end. Call where you arrived at a neighboring node. The average degree of neighboring nodes thus reached approximates the mean branching factor \( \kappa \). In effect, we will take one degree off the average degree found above because the edge we just walked on cannot be used to reach the destination city. We are climbing up a tree, not down (recall how goods find their destination city in section 2.3.6). Also note that the mean branching factor is not just a mean degree \( \langle k \rangle \). We are not hopping from one city to another but climbing a tree from one neighbor to next to reach the destination city. Thus, a city charged with lots of links is more likely to be a neighbor of some city than a poorly connected city, and cities are duly weighted when fed into the mean branching factor. In other words,
Houston is rare while there are quite a few mid-sized cities but that does not mean Houston is hard to reach at random for its rarity. Houston has far more edges than mid-sized cities and we are likely to travel through Houston at some point or another (cf. figure 7(a)). In particular a node of degree \( k \) has a chance proportional to \( kg(k) \) of being at one end of an arbitrary direction on a randomly chosen edge, where \( g(k) \) is a probability density function of (21). Or put differently, if we parachute into a random edge and then flip a coin to decide which direction to go in, we will arrive at a \( k \)-th degree city \( kg(k) \) out of \( \sum_{x=1}^{J} xg(x) \) times. Thus, the mean branching factor is given by (22).

### A.6 Maximum Spacing Estimation

It might be easier to make sense of the use of geometric mean in MSE if we recast it as an analogue of a more familiar, linear regression. The geometric mean of steps here corresponds to ordinary least squares and the arithmetic mean corresponds to a plain sum of residuals. Say we are trying to regress \( y = (-1, 0, 1) \) on \( x = (-1, 0, 1) \). If we aim to minimize the sum of residuals, any real estimate that makes the regression line run through the origin \((0, 0)\) will work, just as much as any estimate will make the arithmetic mean of gaps \(1/J\). We will end up with infinitely many estimates because residual at \( x = 1 \) always offsets the one at \( x = -1 \). To ward off this cancellation problem, we usually try to minimize the sum of squared residuals, which leads to a unique estimate, a 45-degree line. Similarly, the use of geometric mean will solve the indeterminacy problem that comes with arithmetic mean and will promise us sensible estimates.

The geometric mean also comes in handy here. The gap tends to get tighter near the top and/or the bottom of most distributions as the CDF creeps up to one and/or bears down on zero. However, this does not mean New York or New Amsterdam, IN counts less than other cities as a sample. The geometric mean offsets this general tendency and duly stretches small gaps so that these extremities will receive no less attention than the ones in the middle. There is no particular reason to let the mid-sized cities punch above their weight.

On a related matter, we report Kolomogorov-Smirnov (KS) statistic. MSE is similar to KS in
that both KS and the maximand of MSE are a power mean. KS statistic is a power mean of the form
\[
\left\{ \frac{1}{J} \sum_{i} |\text{Empirical } F(s_i) - F(s_i)|^p \right\}^{\frac{1}{p}}
\] (51)

with \( p \to \infty \) (i.e., the maximum of the residuals, the \( L^\infty \) norm), whereas the maximand of MSE is a power mean of the form
\[
\left\{ \frac{1}{J} \sum_{i} (F(s_i) - F(s_{i-1}))^p \right\}^{\frac{1}{p}}
\] (52)

with \( p \to 0 \) (i.e., the geometric mean of the gaps). The way they aggregate the data is where their difference comes in. KS statistic only picks up a single city where the predicted value deviates from the actual value the most. It does not tell us anything about the selected model’s performance over the remainder of cities other than the fact that their gap is tighter than the KS value (but not by how far). On the other hand, the maximand of MSE is determined by the step gap log-averaged over the entire range of the cities, and probably a better measuring tool to gauge the model’s performance in that respect.

To get a sense of what MSE hunts for, consider what happens if we pull out the estimate that minimizes the geometric mean instead. Minimum spacing estimator would dump the entire interval \([0, 1]\) on one particular city \(i\) (any city will do) so that \(F(s_j; \theta) = 0\) for all \(j < i\) and \(F(s_j; \theta) = 1\) for all \(j \geq i\), in which case, the geometric mean would be zero, the smallest value possible (practically the same result when you try to maximize the arithmetic mean as we mentioned above, in the sense that any estimate will be as good as any other). This would make such a pointless estimator. MSE does the exact opposite.

### A.7 Analysis of Variance

Figures 30 and 31 present the kernel estimate of GMP growth rate. Aside from an increase in variance in the lower mid range and decrease in the mid range, there does not seem to be a systematic correlation between GMP and its variance in growth rate.
A.8 Unconditional Scale Economies

To further understand the reasoning behind proposition 3.3.2, conduct comparative statics on $\frac{\beta}{\alpha}$. Imagine that $\beta$ goes down (or equivalently, $\alpha$ goes up). In intracity equilibrium in $P^i$, $B^i(S^i)$ can take any value. In intercity equilibrium in $P$, $B(S^i)$ is subject to utility equalization condition (43). As the expenditure share of housing $\beta$ decreases, we can pack lots of people in a city (because they do not care about the lot size much) and produce lots of composite goods (which they do care about). The story would have ended here in $P^i$, but in $P$, it goes further. This large size makes the city appealing because the marginal rate of substitution of composite goods measured in terms of lot size is getting larger, and people are willing to swap a large parcel of land for only a few composite goods to squeeze Axel in. We still have utility equalization requirement (29) to meet, but people view a large city with small houses more favorably than a small city with large
houses. To offset the rush of people into a large city, the effective wage rate (41) in a city goes down in equilibrium according to (29) (otherwise, utility level in a large city would exceed small city’s). \( P^i \) does not have this autocorrecting mechanism, with which the city steers its population to an increasing returns to scale level. The city size is exogenous in \( P^i \) and \( P^i \) does not factor in the levelling effect of the wage across the cities, but \( P \) does. And because of the existence of the housing market, we do not specifically require \( \delta(S^i) \) to be nonnegative in \( P \). Notice that as \( \beta \) becomes smaller, scale economies also weaken because the residents do not care about housing, and the residents become more willing to scoot over to make room for Axel, even when he does not raise citywide productivity much for the land they have to give up for him. Housing market is indispensable in this sense to observe endogenously induced agglomeration economies in \( P \).

This observation compares to a closed and open monocentric city model (cf. Brueckner [Bru87]). In a closed monocentric city, size is exogenous but utility level is endogenous just as in \( P^i \). In an open monocentric city, size is endogenous but utility level has to match the national level as in \( P \). Since the wage rate depends on city size, \( P \) picks the levelling effect of wage but \( P^i \) does not, and therefore, we have to throw in an additional assumption for \( P^i \).

### A.9 The City-Size Distribution and Gibrat’s Law

Denote discrete time by subscript \( t \) and define \( \Lambda(S^i_t) := a_+(S^i_t) a_-(S^i_t) (S^i_t)^{-\beta} \) and suppose \( \Lambda(\cdot) \) is invertible in the neighborhood of equilibrium \( S^i \). Then \( \Lambda^i_t \Lambda(S^i_t) = K \) from (43) so that \( S^i_t = \Lambda^{-1}(K/\Lambda^i_t) \). With the law of motion \( \Lambda^i_t = (1 + \sigma^i_t)\Lambda^i_{t-1} \), we have

\[
S^i_t = (1 + \epsilon^i_t)S^i_{t-1},
\]

where \( 1 + \epsilon^i_t := 1/\Lambda^{-1}(1 + \sigma^i_t) \). Then

\[
\frac{d \log S^i_t}{dt} \approx \epsilon^i_t,
\]
for a small $\epsilon$, leading to the Gibrat’s law. Computing the city size recursively, (53) also implies

$$\log S_i^t \approx \log S_0^i + \sum_{t=1}^t \epsilon_i^t,$$

which leads to the lognormal city-size distribution as $t \to \infty$ by the central limit theorem. See [Eco04] for details.

REFERENCES


