



# Essays in Economic Geography

## Citation

Ziv, Oren. 2015. Essays in Economic Geography. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:17467205>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

# Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

# Essays in Economic Geography

A dissertation presented

by

Oren D. Ziv

to

The Committee on Higher Degrees in Political Economy and Government

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Political Economy and Government

Harvard University

Cambridge, Massachusetts

May 2015

© 2015 Oren D. Ziv

All rights reserved.

*Dissertation Advisors:*  
**Professor Marc Melitz**  
**Professor Edward Glaeser**

*Author:*  
**Oren D. Ziv**

## **Essays in Economic Geography**

### **Abstract**

While economic geography is concerned chiefly with proximity, models in urban economics eliminate proximity as a relative metric in order to preserve tractability. I introduce a new method of solving spatial models that allows for the consideration of proximity in an economic geography setting while retaining much of the tractability of the urban framework. The first chapter in this thesis introduces the solution method for continuous space geography models and shows how it reduces the complexity of the equilibrium conditions and allows such a model to generate more predictions than was previously possible.

In this chapter, I build a model of firm location decisions in a spatial setting in order to provide a new explanation for the relationship between productivity and density: sorting of heterogeneous firms for market access. This geographic model of sorting breaks observational equivalence between firm sorting and agglomeration forces: under specific conditions, positive shocks to density can negatively affect average productivity through changes in the local composition of firms, inconsistent with models of agglomeration forces without sorting.

Using restricted access establishment-level Census data, I document strong intra-city relationships between location and firm characteristics predicted by the model. I test for evidence of composition effects, instrumenting for the supply of new non-residential real estate construction using the geographic distribution of multi-city real estate developers, and find evidence of firm sorting.

The second chapter in this thesis finds persistent differences in self-reported subjective well-being across U.S. metropolitan areas and uses historical data to show that cities that are now declining were also unhappy in their more prosperous past. The third chapter in this thesis considers the spatial location decisions of multi-unit firms and highlights two previously understudied potential agglomeration and dispersion forces: intra-firm distance costs and market cannibalization.

# Contents

|  |           |
|--|-----------|
| Abstract . . . . .   | iii       |
| Acknowledgments . . . . .  | x         |
| <b>Introduction</b>  | <b>1</b>  |
| <b>1 Density, Productivity, and Sorting</b>                                    | <b>4</b>  |
| 1.1 Introduction . . . . .   | 4         |
| 1.2 Model . . . . .  | 10        |
| 1.2.1 Environment . . . . .  | 11        |
| 1.2.2 Setup . . . . .  | 14        |
| 1.2.3 Equilibrium . . . . .  | 22        |
| 1.3 Predictions of the model . . . . .   | 30        |
| 1.3.1 Predictions for firm and location characteristics . . . . .              | 31        |
| 1.3.2 Composition Effects . . . . .  | 36        |
| 1.4 Evidence . . . . .   | 41        |
| 1.4.1 Data . . . . .   | 42        |
| 1.4.2 Evidence on static relationships . . . . .                               | 43        |
| 1.4.3 Evidence from movers . . . . .   | 48        |
| 1.4.4 Composition Effect . . . . .   | 54        |
| 1.5 Conclusion . . . . .   | 62        |
| <b>2 Unhappy Cities</b>  | <b>67</b> |
| 2.1 Introduction . . . . .   | 67        |
| 2.2 Unhappiness Across Cities . . . . .  | 72        |
| 2.2.1 Are There Significant Differences in Life Satisfaction across Space? . . | 74        |
| 2.2.2 Do Metropolitan Area Differences in Subjective Well-Being Persist? .     | 77        |
| 2.2.3 Is Urbanization Associated with Happiness or Unhappiness? . . . . .      | 79        |
| 2.2.4 Unhappiness and Urban Characteristics . . . . .                          | 80        |
| 2.3 Unhappiness and Urban Decline . . . . .                                    | 84        |
| 2.3.1 Linear Effects of Population Growth and Income . . . . .                 | 84        |

|          |   |            |
|----------|---|------------|
| 2.3.2    | The Non-Linear Relationship between Happiness and Population Growth . . . . .                   | 86         |
| 2.3.3    | Can Urban Disamenities Explain the Correlation between Unhappiness and Urban Decline? . . . . . | 89         |
| 2.3.4    | Is the Unhappiness of Declining Cities New or Old? . . . . .                                    | 95         |
| 2.4      | Why Does Happiness Differ Across Space? . . . . .   | 96         |
| 2.4.1    | Is Happiness a Measure of Utility? . . . . .  | 98         |
| 2.4.2    | Happiness and Utility . . . . .   | 101        |
| 2.5      | Are Individuals Compensated for Unhappiness? . . . . .  | 105        |
| 2.5.1    | Results from Historical and Contemporary Census Data . . . . .                                  | 106        |
| 2.6      | Conclusion . . . . .  | 110        |
| <b>3</b> | <b>The Location Decisions of Multi-establishment Firms in a Spatial and Industrial Network</b>  | <b>112</b> |
| 3.1      | Introduction . . . . .  | 112        |
| 3.2      | Data . . . . .  | 115        |
| 3.3      | Firm productivity and plant dispersion . . . . .  | 116        |
| 3.4      | Evidence on internal economies of density . . . . .   | 118        |
| 3.5      | Market cannibalization . . . . .  | 121        |
| 3.6      | Headquarter services and intra-firm trade . . . . .   | 121        |
| 3.7      | Evidence on external economies of density . . . . .   | 123        |
| 3.8      | Conclusion . . . . .  | 124        |
|          | <b>References</b>   | <b>127</b> |
|          | <b>Appendix A Appendix to Chapter 1</b>   | <b>133</b> |
| A.1      | Proof of sorting equilibrium. . . . .   | 133        |
| A.1.1    | Sorting conditional on non-degenerate distribution . . . . .                                    | 133        |
| A.1.2    | Proof of impossibility of complete non-sorting equilibrium . . . . .                            | 134        |
| A.1.3    | Partial sorting equilibria . . . . .  | 135        |
| A.2      | Landowner density and firm choice decisions . . . . .   | 136        |
| A.3      | Proof of Existence and conditions for uniqueness . . . . .                                      | 138        |
| A.3.1    | Existence . . . . .   | 138        |
| A.3.2    | Uniqueness . . . . .  | 140        |
| A.3.3    | Stability . . . . .   | 140        |
| A.4      | Conditions for a mono-centric equilibrium . . . . .   | 141        |
| A.5      | Incorporating productivity spillovers . . . . .   | 144        |

|                   |  |            |
|-------------------|--|------------|
| <b>Appendix B</b> | <b>Appendix to Chapter 2</b>   | <b>146</b> |
| B.1               | Data Appendix to Glaeser, Gottlieb and Ziv, “Unhappy Cities” . . . . .   | 146        |
| B.1.1             | BRFSS . . . . .  | 146        |
| B.1.2             | Aggregate Data . . . . .   | 151        |
| B.2               | Data on Movers from the National Survey of Families and Households . . . | 151        |
| <b>Appendix C</b> | <b>Appendix to Chapter 3</b>   | <b>154</b> |
| C.1               | Environment . . . . .  | 155        |
| C.1.1             | Agents . . . . .   | 155        |
| C.1.2             | Geography and timing . . . . .   | 155        |
| C.1.3             | Demand for final goods . . . . .   | 155        |
| C.1.4             | Production . . . . .   | 156        |
| C.2               | Worker location decision . . . . .                                       | 159        |
| C.3               | Firm location decisions . . . . .  | 160        |
| C.4               | Equilibrium . . . . .  | 160        |
| C.4.1             | The spatial-industrial network . . . . .                                 | 161        |
| C.4.2             | Labor market clearing . . . . .  | 162        |



## List of Tables

|     |   |     |
|-----|---|-----|
| 1.1 | Summary statistics . . . . .  | 44  |
| 1.2 | Productivity vs establishment density . . . . .   | 45  |
| 1.3 | Other cross-sectional relationships . . . . .   | 46  |
| 1.4 | Sorting on Productivity . . . . .   | 53  |
| 1.5 | Marginal effects of density . . . . .   | 54  |
| 1.6 | Predicted construction expenditures . . . . .   | 56  |
| 1.7 | Composition effects at urban core . . . . .   | 59  |
| 1.8 | Composition effects at neighboring tracts . . . . .   | 61  |
|     |   |     |
| 2.1 | Happiness levels across space, BRFSS . . . . .  | 82  |
| 2.2 | Happiness and urban characteristics, BRFSS . . . . .  | 85  |
| 2.3 | Happiness and urban population growth differences, BRFSS . . . . .                          | 88  |
| 2.4 | Happiness, urban decline, and mobility, NSFH . . . . .                                      | 91  |
| 2.5 | Happiness regressions using the General Social Survey . . . . .                             | 94  |
| 2.6 | Income, housing costs, population growth, and happiness, 1940 to 2000<br>Censuses . . . . . | 107 |
|     |   |     |
| 3.1 | Summary statistics . . . . .  | 115 |
| 3.2 | Effect of distance from firm center on establishment productivity . . . . .                 | 119 |
| 3.3 | Market cannibalization . . . . .  | 122 |
| 3.4 | Internal agglomeration economies . . . . .  | 123 |
| 3.5 | External agglomeration economies . . . . .  | 124 |

## List of Figures

|     |  |     |
|-----|--|-----|
| 1.1 | Productivity vs Density . . . . .  | 5   |
| 1.2 | Productivity vs Density, city by industry by year fixed effects . . . . .                            | 5   |
| 1.3 | Log output per worker vs log establishment density . . . . .   | 47  |
| 1.4 | Log value added per worker vs log establishment density . . . . .                                    | 48  |
| 1.5 | Log establishment size vs log establishment density . . . . .  | 49  |
| 1.6 | Log output per worker vs log rent per worker . . . . .   | 50  |
| 1.7 | Log output per worker vs log population density . . . . .  | 51  |
| 1.8 | Log gross margin per worker vs log establishment . . . . .   | 52  |
| 2.1 | Estimated Metropolitan and Rural Area Adjusted Happiness . . . . .                                   | 76  |
| 2.2 | Urban-rural happiness gradient vs GDP across countries . . . . .                                     | 80  |
| 2.3 | Estimated Metropolitan and Rural Area Adjusted Happiness . . . . .                                   | 86  |
| 2.4 | MSA rent vs adjusted happiness . . . . .   | 109 |
| 3.1 | Log average output per worker vs log average distance from firm center . . .                         | 116 |
| 3.2 | Log average value added per worker vs log average distance from firm center                          | 117 |
| 3.3 | Log average employment per worker vs log average distance from firm center                           | 117 |
| 3.4 | Log average output per worker vs log average establishment density . . . . .                         | 118 |
| 3.5 | Log output per worker vs log distance from firm center, firm-year and age<br>fixed effects . . . . . | 120 |

## Acknowledgments

I am forever indebted to my thesis advisors, Pol Antràs, Ed Glaeser, Elhanan Helpman, and Marc Melitz, without whom this work would not have been possible. I am also grateful to James Anderson, Jim Davis, Cecile Gaubert, Wayne Gray, Larry Katz, Naomi Hausman, Ben Li, and Esteban Rossi-Hansberg, and a decade of advice, support, and friendship from Ned Phelps. I am also indebted to David Rezza Baqaee, Kirill Borusyak, Andrew Garin, Jamie Lee, Benjamin Schoefer, and seminar participants at Harvard University and Boston College. I acknowledge support for this project from the NSF via grant numbers DGE0644491 and DGE1144152, the NE-UTC, the Taubman Center for State and Local Government, and the Harvard University Program on Inequality and Social Policy. All mistakes are my own. Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.

To my parents, Elad, Ohad, Etay, Tessa, and Numi.

# Introduction

Two distinct literatures in economics have attempted to consider the causes and consequences of the dispersion of economic activity across space: urban economics and economic geography. While similar in subject, these literatures differ immensely in method. Economic geography is concerned chiefly with the economics of proximity, modeling the complex spatial interconnections between points and the effect of this spatial network on the decisions of agents, as they choose locations in the network and how to interact with other nodes. The measure of this network is distance.

From the perspective of economic geography, urban economics deals with cities as though they were islands: trading freely (with no transportation costs) in certain industries while not at all in others (such as the real estate market). In this framework, proximity as a relative metric is fundamentally eliminated. Instead, cities are characterized only by their own qualities. In urban economics, the difference between New York and Philadelphia has everything to do with their relative sizes but nothing to do with their relative proximity to Boston.

In return for eschewing the consideration of these complex interdependencies, urban economic models gain tractability. The goal of this thesis is to change the calculus surrounding this tradeoff. I introduce a new method of solving spatial models that allows for the consideration of proximity in an economic geography setting while retaining much of the tractability of the urban framework.

The second chapter, co-written with Edward Glaeser and Joshua Gottlieb, investigates spatial patterns in subjective well-being. We observe that there are persistent differences in

self-reported subjective well-being across U.S. metropolitan areas, and residents of declining cities appear less happy than others. Yet some people continue to move to these areas, and newer residents appear to be as unhappy as longer term residents. While historical data on happiness are limited, the available facts suggest that cities that are now declining were also unhappy in their more prosperous past. These facts support the view that individuals don't maximize happiness alone, but include it in the utility function along with other arguments. People may trade off happiness against other competing objectives.

The third chapter in this thesis is an essay in economic geography. Multi-establishment firms face a unique set of tradeoffs when choosing establishment locations: establishments placed close together can cannibalize each other's market, while spacing plants far apart can create supply chain and management problems.

These forces are responsible for the unique patterns of multi-establishment firm geography. This paper documents these stylized facts: more productive firms cover larger distances and are less densely populated by establishments; firms grow out radially from their centers; new establishments cannibalize sales of nearby existing establishments; and plants that are further away from the firm's center experience productivity losses.

I adapt the Tintelnot (2014) model of export platforms build a model of multi-establishment firm geography. Firms trade off firm organizational advantage and cannibalization effects along with market access and traditional agglomeration forces (proximity to suppliers, productivity spillovers, and natural advantages). Introductory chapter that talks about all three papers for a little bit longer than the abstract.

It is my first chapter which bridges these two literatures. In it, I begin with the old observation that firms in denser locations are more productive. This persistent relationship has been used as evidence for the existence of agglomeration forces such as productivity spillovers.

This paper presents an alternative hypothesis: In an economic geography model where firms choose locations to be close to their markets, sorting endogenously generates the density-productivity relationship in the absence of any pecuniary or non-pecuniary produc-

tivity spillovers. I introduce a new solution method for continuous space geography models that dramatically reduces the complexity of the equilibrium conditions and allows such model to generate more predictions than was previously possible in realistic geographic networks, where firms and workers choose their locations and where these choices have material consequences on the distribution of economic activity.

Other cross-sectional relationships that have traditionally been used as evidence of agglomeration forces can also be derived in this environment where no such productivity spillovers exists. This geographic model of sorting breaks observational equivalence between firm sorting and agglomeration forces and allows for an indirect test of firm sorting. Under specific conditions, positive shocks to density can negatively affect average productivity through changes in the local composition of firms, inconsistent with models of agglomeration forces without sorting.

Using restricted access establishment-level Census data, I document strong intra-city relationships between location and firm characteristics predicted by the model. I use the data to test for evidence of composition effects, instrumenting for the supply of new non-residential real estate construction using the geographic distribution of multi-city real estate developers, and find evidence of firm sorting.

# Chapter 1

## Density, Productivity, and Sorting

### 1.1 Introduction

Across and within cities, firms in dense locations – measured by population, establishment, or employment density – are more productive.<sup>1</sup> Figures 1.1 and 1.2 plot the relationship between total factor productivity<sup>2</sup> and establishment density for a sample of US manufacturing firms. Because transport costs are thought to be low, previous work has concluded that differences in access to local markets alone could not account for this persistent relationship. Because of this, the relationship has been used as evidence for the existence of productivity spillovers that work either through learning or inter-industry linkages.<sup>3</sup> These theories collectively posit a particular direction of causation: otherwise homogeneous firms become more productive when locating in denser areas.

This paper offers an alternative hypothesis: when firm mobility is added to a model

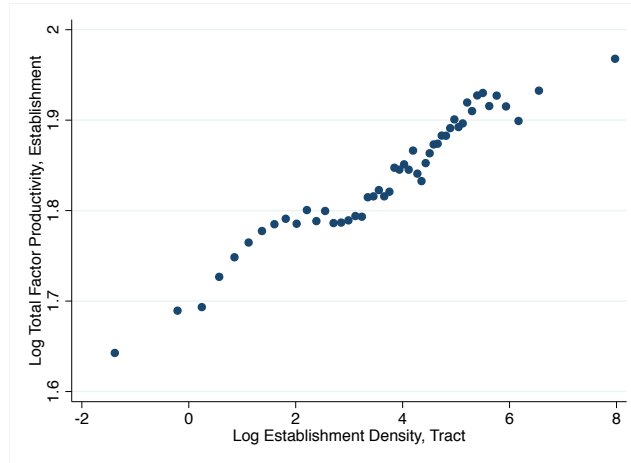
---

<sup>1</sup>This stylized fact has been validated in multiple periods and across multiple continents. Ciccone and Hall (1996) were the first to differentiate between density and city size with respect to productivity. Recently work by Maré *et al.* (2006) and Combes *et al.* (2010) confirm the relationship using data from New Zealand and France, respectively.

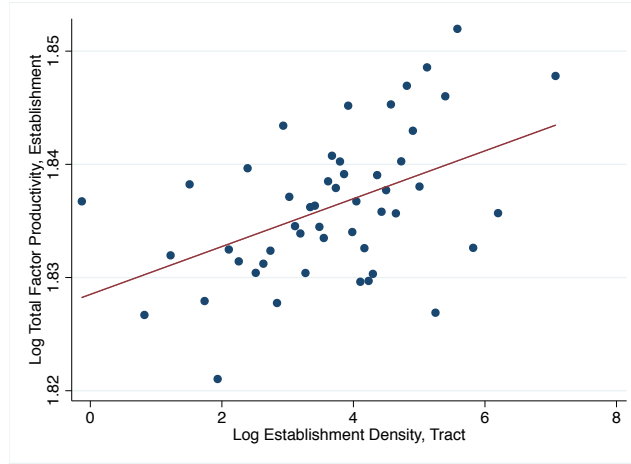
<sup>2</sup>TFP measures at the establishment level for all responders to the Annual Survey of Manufacturers, from Petrin *et al.* (2011).

<sup>3</sup>See (Rosenthal and Strange, 2004b).





**Figure 1.1:** *Productivity vs Density*



**Figure 1.2:** *Productivity vs Density, city by industry by year fixed effects*

with monopolistic competition and heterogeneous firms,<sup>4</sup> sorting for market access generates productivity and density differences across locations, and endogenously generates the relationship between the two – even in the absence of pecuniary or non-pecuniary productivity spillovers and when underlying differences in demand across locations are small. In Section 2, I present a model in which locations differ by transport costs to and from markets at all other locations, affecting demand at each location. More productive entrepreneurs outbid others for locations with higher market potential, trading fixed costs

---

<sup>4</sup>Melitz (2003)

of higher rents for higher variable profit. In turn, these firms attract more workers to those locations. This process generates equilibrium differences in market access, real estate prices, and firm productivity. Higher prices induce landowners to provide more density in areas where more productive firms locate. Firm sorting amplifies the same centripetal forces present in the new-economic geography literature.<sup>5</sup> Moreover, when firms sort on market access, the relationship between density and productivity arises endogenously.

The model is defined over a broad set of domains so that it may be taken to data from actual urban geographies with minimal analogy. This flexibility is similar to Allen and Arkolakis (2013), where trade takes place in a Ricardian framework with immobile, homogenous firms. My model adds a production setting with mobile, heterogeneous firms where the density of firms at each location is endogenously determined by real estate prices.

Differences between locations are a function of access to markets and variable costs, and therefore a function of the location decisions of all other firms and workers as well as of the space provision decisions of landowners. This complexity is endemic to new economic geography models, where it is often impossible to guarantee uniqueness or derive analytic relationships between key observables under realistic conditions.<sup>6</sup>

I introduce a novel solution method for this class of models, using location advantage as a sufficient statistic for the economic activity at a given location, an endogenously determined index,<sup>7</sup> and solving the model in two steps. First, I solve the firm, worker, and landowner decisions with respect to the index, and then use the general equilibrium conditions to solve for the mapping of index values onto locations. The first step reduces the dimensionality of these problems. The latter step recovers the complexity of the geographic interconnections and is necessary in order to understand the relative strength of the geographic centripetal and centrifugal forces or to describe the effects of geography-based policies. This second step

---

<sup>5</sup>See Fujita *et al.* (1999)

<sup>6</sup>Fujita *et al.* (1999); Rossi-Hansberg (2005); Allen and Arkolakis (2013); Kyriakopoulou and Xepapadeas (2013)

<sup>7</sup>Davis and Dingel (2013) use a system of cities approach where distance from the city center is exogenously given as an index for location advantage. By contrast, here the index is endogenously determined across all locations, arising out of the effects of geographic frictions on the firms' production function.

does not deliver a closed-form solution and the equilibrium may not be unique, rendering it uninformative. However, the equilibrium conditions derived with respect to the index allow me to derive predictions based on observables that hold in any equilibrium. In Appendix 5, I show how this method can be applied to geography models with a broader set of assumptions, including location-specific and geographically determined spillovers.

Section 3 uses the model to derive predictions. I derive analytic relationships between observables (firm or location characteristics) and the location index, and use these in turn to derive analytic relationships of observables with respect to each other. In this way, the model delivers relationships between firm productivity, size, and profits, and firm density, employment or population density, and rents.

This paper also makes a theoretical and empirical contribution to a recent literature that proposes both the sorting<sup>8</sup> and selection<sup>9</sup> hypotheses to explain the closely related city-size-productivity relationship. In these models, firms are attracted to and sort into large cities in order to gain access to city-level agglomeration forces derived from city size. Gaubert (2014) assumes super-modularity directly between firm productivity and city size. In Behrens *et al.* (2010), productive firms take advantage of larger bases of local intermediates in larger cities. When, as in this literature, both forces are posited in tandem, the two become empirically indistinguishable. Any shock to the size of a city also affects the quality of firms sorting into the city. This fundamental identification problem, termed “observational equivalence” by Ellison and Glaeser (1997), prohibits the disentangling of firm and location characteristics when location decisions are endogenous.

The introduction of geography breaks the observational equivalence result. In my model, the sorting-induced density-productivity relationship occurs even in the absence of agglomeration forces. Firms sort on location advantage, which is based on networked proximity to other markets. Higher real estate prices in advantageous locations increase density. In equilibrium, advantageous locations are denser, but a location’s density does not

---

<sup>8</sup>Gaubert (2014); Behrens *et al.* (2012); Maré and Graham (2009)

<sup>9</sup>Baldwin and Okubo (2006); Nocke (2006); Combes *et al.* (2012)

determine its advantage through agglomeration forces. Rather, density and firm quality at each location endogenously respond to geographic proximity.

This last theoretical innovation provides the key empirical advantage that enables me to distinguish firm sorting from agglomeration forces. Because density and firm quality both respond endogenously to location advantage, shocks to the sorting pattern of firms can affect the quality of firms at a location by changing the composition of firms the location, without affecting the underlying determinants of location advantage. While the model sometimes predicts positive relationships between positive shocks to density and average productivity, as predicted in an agglomeration force model, negative relationships emerge under specific circumstances. These negative relationships are not natural predictions of agglomeration force models. Introducing geography reframes the city-size-productivity literature in terms of the density-productivity relationship and provides this new mechanism underlying the sorting pattern of firms which can be tested. I test for and find evidence of composition effects that are consistent with a model where sorting and density are both endogenously responsive to location advantage.

Section 4 tests the predictions of the model using data collected from US Economic Censuses and Surveys between 1992 and 2007. The data show that firms in locations with higher establishment density have more sales and employees, are more productive, and pay higher rent and more rent per worker. Population density and productivity are positively related. These relationships hold within cities and across the US.

However, these predictions can also be derived from models of agglomeration forces. To begin testing for sorting, I first examine the location decisions of firms that expand and firms that relocate plants, using previous period productivity to predict new location density, and the subsequent effects of productivity five years forward. Although this strategy can't rule out sorting on unobservables linked to density, I find evidence for inter- and intra-city sorting as well as mixed evidence for density effects.

Finally, I test the sorting hypothesis using the composition effects. To do this, I must isolate exogenous shocks to the supply of density. I use data from the Census of Finance and

Insurance on the geography and construction expenditures of real estate development firms. Because commercial real estate development requires liquid collateral,<sup>10</sup> such firms are exposed to real estate shocks in multiple cities, and price shocks in one city can both affect firm-level collateral and transfer resources away from projects in relatively lower-shocked cities. Both these channels appear to be in effect in the data.

I test two different composition effects. First, I isolate what the model predicts as the highest index locations in each city, ranking each tract in the city and dividing each city into percentiles. Construction of new non-residential real estate within the highest density-percentiles results in lower-productivity entrants. Second, I show that construction in a given tract lowers the productivity of entrants in relatively lower-rank tracts, but has no effect on higher-rank tracts.

These findings are consistent with the sorting hypothesis and inconsistent with the baseline agglomeration hypothesis, where higher density increases firm productivity. To be sure, these results cannot be interpreted as a rejection of agglomeration forces; models including both sorting and agglomeration forces may also predict the negative relationship I find. However, these results represent the first affirmation of the existence of intra-city firm sorting, and demonstrate the empirical flexibility of the indexing strategy.

My model has significant ramifications for a number of urban policies. As in other models of sorting, the model implies that significant mismeasurement of agglomeration forces may over or understate the benefits of place-based urban policies, where municipalities subsidize the relocation of large, productive firms with the implicit assumption that incumbents gain pecuniary or non-pecuniary externalities. Moreover, the model has counter-intuitive predictions for the effects of relaxing zoning laws. In models with agglomeration forces, it is natural to think that allowing for the accumulation of more density will increase productivity locally. In this model, the opposite is often true, as new entrants can be of lower quality and reduce average productivity. This prediction stands in stark contrast to accepted wisdom.

---

<sup>10</sup>Gyourko (2009) reports 1-to-1 leverage ratios.

Furthermore, my findings have important ramifications for our understanding of the gradient of land prices in cities. In models with agglomeration forces, firms pay for the productive amenities of cities, and a hedonic calculation would recover the full value of a location amenity. In this model, over and above differences in location quality, rents in central areas are bid up by more productive firms, and the overall rent gradient reflects both underlying differences in location quality, and differences in the underlying distribution in firm quality: if we want to understand why the centers of cities are more expensive than their peripheries, we must understand the sorting behavior of firms within cities.

## 1.2 Model

In this section, I outline the environment of the model, including assumptions on the geography and transportation costs in the domain. I set out the optimization problems for each of the three agents: workers, landowners, and entrepreneurs.

I then introduce the change of variable that will act as an index of location advantage. I separate the potential variable profit at each location into location-specific and firm-specific terms. The location-specific terms can be decomposed into prices and market size at all other locations, and the geographic relationship between locations. Together, these endogenously determined terms constitute all the geographic terms that affect variable profits, and the location index is the specific functional form for the effects of geography on variable profits. Effectively, a location's advantage in this model is its market access. In Appendix 5, I show how a broader set of production functions that include geographic productivity spillovers can be incorporated into this framework.

My solution method first solves the landowner space provision and firm location decisions with respect to this index. Because the terms comprising the index are geographic, their values cannot be determined in complicated spaces where closed-form solutions do not exist. However, the index is a sufficient statistic for all the effects of geography on firm worker and landowner decisions, so the mapping of firms, workers, and density to the index characterizes the full set of potential equilibria. To my knowledge, this is the first paper to

employ such a method.

The full benefit of this solution method will be made more clear here and in the following section where I explore the model's predictions. Because the first order conditions of the agents are determined with respect to the index, they are free of geographic variables and hold under any equilibrium distribution of firms and workers. The predictions of the model will be derived from these first order conditions and therefore are also true in any equilibrium.

That said, to evaluate a particular equilibrium, it becomes necessary to reintroduce geographic variables. For instance, to evaluate the relative strengths of the geographic forces comprising location advantage, or to evaluate the effects of decreased trade costs due to transportation infrastructure development, the model must be solved for a particular equilibrium. To do this, I must map locations to index values.

The second step of the solution method is to solve for the mapping of location advantage to locations. I derive the equilibrium conditions for this mapping, which pins down both prices and market access in any particular geography. This system of nonlinear integral equations resembles the equilibrium conditions in Allen and Arkolakis (2013) as well as many other economic geography models. I show that under the particular assumptions of this model, an equilibrium must always exist. Uniqueness is guaranteed under particular conditions for transportation costs that are given in Appendix 3. Because these conditions may not be met, and because the imposition of uniqueness generates no further predictions in this context, I do not impose them.

### **1.2.1 Environment**

The following subsection describes the geographic and economic environment of the model.

#### **Geography**

The goal of this model will be to be able to accept as a domain any realistic geography for which we have data. As such, the model is defined over  $S$ , any compact subset of Euclidean

space  $R^n$ ,  $n \in \mathbb{N}$ . All economic activities, production and consumption, take place at points  $i \in S$ . The flexibility of this domain allows for applications to real geographies, such as two-dimensional planes. The compact nature of the space guarantees the existence of a boundary, locations that are relatively distant. Given further assumptions on the transportation costs stated below, locations on the boundary will inevitably be economically remote, and thus less advantageous. This will be crucial in ensuring at least partial sorting.

Three kinds of agents take part in production and consumption in the space: landowners, workers, and entrepreneurs. The space is filled with a mass of immobile landowners, each endowed with a point  $i \in S$ . I refer to this as a landowner's location. All locations have landowners. Landowner locations are fixed. Workers and entrepreneurs choose their location (and their landowner). The requirement that economic activity is assigned a location constitutes the fundamental friction posed by space.

The space is further defined by a function governing transportation costs between all points in the space. Goods are sold from one point  $i$  to another point  $j$  with continuous, differentiable, and symmetric iceberg transport costs

$$1 < \tau(i, j) < \infty.$$

To help satisfy the existence of a sorting equilibrium, I impose symmetry and the triangle inequality on  $\tau(i, j)$ .<sup>11</sup> No other functional form is placed on  $\tau$ . There's no direct comparison that can be made between transportation costs between  $i$  and any two other locations  $j_1$  and  $j_2$ .  $\tau(i, j_1)$  and  $\tau(i, j_2)$ , even if we do know something about the Euclidian distance between  $i$  and the other points, e.g. if the vector  $\vec{ij_1} > \vec{ij_2}$ .

Locations differ by their relative proximity to other locations. On its own, this exogenous geography will not drive the location decisions of firms and workers. Rather, a location's

---

<sup>11</sup>This is a technical assumption that will be necessary in order to ensure that points on the boundary are actually more distant from all other locations than nearby points, so that market access and prices both improve away from the boundary. A less strict condition is sufficient. For any point  $k$  on the vector  $\vec{ij}$ ,  $\tau(i, j) > \tau(i, k)$  and  $\tau(i, j) > \tau(k, j)$ ; intuitively for any journey between two points  $i$  and  $j$ , transportation costs are lower for stops along the way. If travel between  $i$  and  $j$  always takes place on the euclidean vector  $\vec{ij}$ , this constraint intuitively means shorter trips facing the same geography must be less costly. In a world where shortest cost trips may include circuitous routes, this constraint may be unrealistic.



advantage is determined by the equilibrium decision of workers, firms, and landowners. The joint actions of exogenous spatial characteristics and the endogenous economic potential available to agents at each location drive the sorting behavior of firms and the predictions of the model.

## Production

A fixed set of firms use labor to produce differentiated goods. Firms are required to locate somewhere in order to produce. In addition, firms employ labor at their locations and pay location-specific wages. The total market of available labor is equal to some mass of  $L$  workers. Differentiated goods are sold to consumers (workers, entrepreneurs, and landowners) across the entire space. Access to consumers, local production costs, and rents drive the location decisions of firms.

Landowners produce units of non-residential space and provide that space to a density of firms.

## Consumption

All three agent types purchase and consume a CES aggregate of all the differentiated goods produced at all locations with an elasticity of substitution  $\sigma > 1$ .<sup>12,13</sup> The quantity of each good consumed will vary in equilibrium by location. Consumers will substitute towards goods produced locally, as such goods face lower transportation costs, and goods produced

---

<sup>12</sup>Fixing landowners consumption at their location assumes balanced trade. This assumption simplifies the notation significantly but does not drive the results. Instead, landowners and entrepreneurs can be assumed to be companies in which each worker owns stock. It is crucial, however, that some consumption remain localized. If all three agents could separate production and consumption choices entirely, then geographic frictions would no longer be a defining element of this model.

<sup>13</sup>Specifically, all three agents maximize their utility

$$U = \left[ \int_{\omega \in \Omega} q(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right]^{\frac{\sigma}{\sigma-1}}$$

where  $\sigma > 1$  is the elasticity of substitution and  $\omega \in \Omega$  is a good in the set of all available goods  $\Omega$ .

by more productive firms, as the factory prices of such goods will be lower.<sup>14</sup>

## Timing

The model is static and all production, location, and consumption decisions are made simultaneously by all agents.

### 1.2.2 Setup

#### Worker location decision

A mass  $L$  of homogenous workers provide one unit of labor inelastically. Workers consume at their location. Because my focus is on the commercial real estate market and the location decisions of firms, I will assume that workers do not participate in a real estate market.<sup>15</sup> Because all goods will be sold to all locations, workers view locations as varying only by their respective price indexes. For any two locations, homogeneous workers must be indifferent between those two locations. In equilibrium, wages must exactly offset prices such that worker utility equalizes across space.

---

<sup>14</sup>The local price index at location  $i$  is defined as

$$P(i)^{1-\sigma} = \int_{\omega \in \Omega} p(\omega, i)^{1-\sigma} d\omega$$

where  $p(\omega, i)$  is the location-specific price of each good.

<sup>15</sup>An alternative would model workers as having an inelastic per capital housing demand at their location and the residential real estate market, segregated from the commercial real estate market, as a competitive fringe of builders with a constant marginal cost denominated in local labor. The worker equilibrium condition would remain similar to the one below, effectively multiplying the current location-specific wage by a constant equal to one plus the marginal cost of residential real estate. A more robust version would allow landowners to supply both residential and commercial real estate, and workers to commute between locations at a cost. As in Lucas and Rossi-Hansberg (2002) or Ahlfelt et. al. (2012), the relative density of commercial and residential activity would vary across locations. In this scenario, employment, but not firm density, is always positively correlated with firm productivity.

## Firm location and pricing decisions

A set of entrepreneurs receive heterogeneous productivities and will each create a single firm to sell final goods to all agents at all locations.<sup>16</sup> Entrepreneurs face three decisions: (1) whether to produce (2) where to locate and (3) how to price their firm's good. Entrepreneurs simultaneously solve these three problems. An entrepreneur's location decision affects the variable and fixed costs of production they face as well as the demand they will face. The entrepreneur's optimal choice can be found by solving these three decisions in reverse: first determining the optimal price of the good at each potential location, then the optimal location given the pricing rule at each location, and finally whether to produce given the profits yielded by the optimal location.

Entrepreneurs draw a productivity  $\psi$  from some distribution  $G(\hat{\psi})$  where  $\hat{\psi} = \psi^{\frac{1}{\sigma-1}}$ .<sup>17</sup> The distribution need only be assumed to have an upper and lower bound,  $0 < \psi_L < \psi_U < \infty$ . The bounds on the distribution are technical assumptions that help ensure the existence of an equilibrium.

This firm-specific productivity lowers the marginal cost of production which is a function of a location-specific wage:

$$MC = \frac{w(i)}{\hat{\psi}}.$$

Note that in equilibrium, real wages will be identical across locations, thus the real marginal costs for a specific firm will not vary across locations: there are no location productivity advantages here.

Following the literature, at any location, the firm's price decision will be a constant markup over its marginal cost.<sup>18</sup> Through the wage, potential marginal costs, and therefore

---

<sup>16</sup>I will use entrepreneurs and firms interchangeably throughout.

<sup>17</sup>Defining the distribution over a function of the productivity parameter rather than the productivity parameter itself simplifies the algebra.

<sup>18</sup>For a firm with a given productivity parameter  $\psi$  at a given location  $i$ , the optimal factory price will be

$$p(\psi, i) = \frac{w(i)}{\rho \cdot \hat{\psi}}$$

the final goods' prices, will vary by location; the pricing decision can be folded into the location decision.

Because the wage and transportation costs to other markets differ across locations, firms face higher or lower demand at different locations. Consumers substitute towards local firms, whose goods pay lower transport costs, and firms at locations with lower marginal costs. Because of this variation in demand across locations, firms reap different levels of variable profit in different locations.

The firm pays a fixed cost of rent  $\phi(i)$  (which is denominated in terms of units of consumption) to the landowner in order to rent space at location  $i$ . Rent does not depend on firm size.

Firms face a tradeoff between higher real variable profits and higher real rents. Entrepreneurs' utility is maximized when their real incomes, their firms' profits (in terms of price-index bundles of final goods), are maximized. The firm's maximization function can therefore be written as

$$i^* = \operatorname{argmax}_{i \in S} \{\pi_f(i)\} = \operatorname{argmax}_{i \in S} \{r(\psi, i)/\sigma - \phi(i)\}$$

where  $r(\psi, i)$  is the revenue of a firm with productivity parameter  $\psi$  at location  $i$ .

In order to examine the variable profit  $r(\psi, i)/\sigma$ , I first present the revenue, in units of consumption at  $i$ , from selling from point  $i$  to point  $j$ :

$$r_j(\psi, i) = \frac{\left[ \frac{w(i) \cdot \tau(i, j)}{P(j) \rho \hat{\psi}} \right]^{1-\sigma} R(j)}{P(i)} .$$

Revenue is a function of the markup, the firm's productivity  $\psi$ , the wage at  $i$ , as well as the price index at  $j$ ,  $P(j)$ , the nominal size of the market at  $j$ ,  $R(j)$ , the iceberg transportation costs between  $i$  and  $j$ ,  $\tau(i, j)$ , and the price index at  $i$ ,  $P(i)$ . Summing over all markets  $j \in S$ , the expression for the variable profit at point  $i$ :

$$r(\psi, i)/\sigma = \int_{j \in S} \frac{w(i)^{1-\sigma} \cdot \tau(i, j)^{1-\sigma} R(j)}{P(i) P(j)^{1-\sigma} \rho^{1-\sigma} \psi} \frac{1}{\sigma} dj$$

---

where  $\rho = \frac{\sigma}{\sigma-1}$  is one over the optimal markup.

The above equation yields the relationship between a firm's variable profit and its location. Rearranging terms, it is possible to separate the location-specific effects on variable profit from firm-specific productivity:

$$r(\psi, i)/\sigma = \psi \cdot \eta(i) \quad (1.1)$$

where  $\eta(i)$  is the location-specific advantage term defined as

$$\eta(i) = \int_{j \in S} \frac{w(i)^{1-\sigma} \tau(i, j)^{1-\sigma} R(j)}{\sigma \rho^{1-\sigma} P(i) P(j)^{1-\sigma}} dj.$$

The value of  $\eta(i)$  is dependent on the variable cost of production at  $i$ ,  $w(i)$ , and the transportation cost-weighted proximity to markets  $j$ , both a function of the market size of  $j$  and the price index at  $j$ . The market potential at  $i$  can be expressed as:

$$\text{Market Potential} = \frac{\eta(i) P(i)}{w(i)^{1-\sigma}} = \int_{j \in S} \left[ \frac{\tau(i, j)}{\rho P(j)} \right]^{1-\sigma} \frac{R(j)}{\sigma} dj.$$

This equation gives some intuition for the centripetal and centrifugal forces governing the model. The parameters of the utility function, the equilibrium distribution of firms and workers, and the exogenous geography (through  $\tau$ ) jointly determine a location's market access. Locations that are relatively proximate to more productive firms and far from less productive firms will have lower price indexes, which decrease the effective market for a given firm. On the other hand, larger firms, their workers and landowners together form larger markets, and relative proximity to these markets increases market access via  $R(j)$ . Together with the variable costs at a given location  $i$ , these forces govern location  $i$ 's access to markets.

In addition to effects via market access, proximity to other firms affects location advantage at  $i$  via  $P(i)$ , the price index at  $i$ . Proximity to productive firms reduces the price index at  $i$ , which reduces the wage  $w(i)$ , and thus marginal cost at  $i$ , and directly increases entrepreneur utility by increasing the total amount of consumption for a given level of profits.

Taken together, proximity to high-productivity firms increases affects advantage by (1) increasing the size of local markets, (2) stiffening competition, through lower price indexes in nearby locations, (3) decreasing marginal cost of production, and (4) increasing entrepreneur utility. The first two channels operate through market access, while the last two operate through effects on the price index at location  $i$ .

In equation (2), the ability to separate location- and firm-specific contribution to variable profit is a result of the particular assumptions of the production function. In particular, two features of equation (2) are crucial.

First, higher  $\eta$  locations have higher variable profits for firms holding  $\psi$  constant. All firms agree on which locations yield the highest variable profit. This ranking orders every location  $i \in S$  according to  $\eta(i)$ . For any two locations  $i$  and  $j$ , for any given level of firm productivity, marginal profits will be higher where  $\eta$  is higher. The first proposition restates this:

*Lemma 1* : For any two locations  $\eta_1, \eta_2 \in [\bar{\eta}, \underline{\eta}]$ , if  $\eta_1 > \eta_2$ , then  $r(\psi, \eta_1) > r(\psi, \eta_2)$  i.e., variable profits at  $\eta_1$  are higher for all firms.

Lemma 1 follows directly from the production function and definition of  $\eta$ .

Second,  $\hat{\psi}$  enters multiplicatively with  $\eta$ . As a consequence, difference in variable profit between any two locations is higher for more productive firms.

*Lemma 2* : For any  $\eta_1, \eta_2 \in [\bar{\eta}, \underline{\eta}]$ , if  $\eta_1 > \eta_2$ , and  $\psi_1 > \psi_2$ , then  $r(\psi_1, \eta_1) - r(\psi_1, \eta_2) > r(\psi_2, \eta_1) - r(\psi_2, \eta_2)$ .

Lemma 2 also follows directly from equation (2) and is a result of the super-modularity between location and productivity assumed in the production function. Locations that have higher equilibrium market access will be more sought after by more productive firms. As in the assignment literature,<sup>19</sup> this condition will, in equilibrium, lead firms to sort on productivity, with more productive firms capturing locations with greater market access.

---

<sup>19</sup>Costinot and Vogel (2009) use log-supermodularity to ensure matching.

The equilibrium parameter  $\eta$  and the ability to rank any  $i \in S$  according to  $\eta$  will be central to the subsequent analysis. It is important at this stage to reiterate that the mapping of  $\eta$  into locations is defined endogenously; it will be dependent on a particular arrangement of firms and workers in the space. As the analysis above demonstrates, the equilibrium decision of each firm depends on the decisions of all other firms. The complexity of this problem is enormous. Prices and market access at any given location  $i$  will be impossible to pin down without knowing the prices and market access at every other location.

The introduction of the change of variable,  $\eta$ , greatly reduces the complexity of the task at hand. Without knowing the  $\eta$ -value of a particular location, the subsequent analysis will make claims on the economic activity at a given location  $i$  conditional on the value of  $\eta$  at  $i$ . After solving for the equilibrium behaviors of firms, landowner and workers with respect to  $\eta$ , it will further be necessary to show the mapping  $S \rightarrow [\underline{\eta}, \bar{\eta}]$  must exist. While this means predictions cannot be made regarding the economic activity at a specific location, I will subsequently show that the introduction of the endogenous variable allows for the derivation of analytic relationship between key economic variables in any equilibrium.

Finally, firms produce if their profits, conditional on their optimal location decision, are above zero:

$$r(\psi, \eta^*)/\sigma \geq \phi^*$$

where  $\eta^*$  is the  $\eta$  of the profit maximizing location for firm with productivity  $\psi$ , and  $\phi^*$  is the rent at that optimal location.

### **Landowner supply decision**

Each landowner is endowed with a location and must decide how much density to provide to firms at that location. Density is provided at increasing marginal construction cost according to an invertible, twice-differentiable cost function  $c(h)$ :

$$c'(h), c''(h) > 0, c(0) = 0$$

where  $h$  is the density of firms at a particular location. Construction costs are denominated in baskets of final goods and no labor is required in construction.<sup>20</sup>

Landowners tradeoff the price-adjusted costs of providing density  $c(h)$  against the price-adjusted rents  $\phi$ . In equilibrium, the landowner chooses rent  $\phi$  and density  $h$  to maximize profits

$$\pi_l = h\phi - c(h).$$

From the above equation, holding firm density,  $h$  constant, higher rents increase profits unconditionally. The landowner's choice of  $h$  is unconstrained. However, the choice of  $\phi$  is constrained by the participation constraints of the firms: given each firm's outside option, the landowners' pricing decision will affect the types of firms willing to locate at her location, and she must take this into account when setting rents.

To simplify this problem, note that for any given rent charged, the remaining choice of  $h$  can be expressed as

$$\tilde{\pi}_l(\phi) = \max_h \{h\phi - c(h)\}$$

Given any rent  $\phi$ , optimal density will set marginal costs of density provision equal to marginal revenue,  $\phi$ .<sup>21</sup>

By the envelope theorem, the landowner's profits, conditional on optimally chosen density, are increasing in  $\phi$

$$\tilde{\pi}'_l(\phi) = h > 0.<sup>22</sup>$$

---

<sup>20</sup>Real construction costs differ from location to location due to differences in the price index. Landowners at higher  $P(i)$  locations face higher unit costs of building, as real wages are higher at such locations. However, they also receive higher real rents (controlling for  $\eta$  since  $\phi(\eta)$  is denominated in units of final goods and therefore real rents are higher, all else equal, at locations with higher  $P(i)$  as well. Therefore I write the cost function  $c(h)$  as denominated in units of final goods. Thus, the landowners' provision decision is homogenous degree zero with respect to  $P(i)$ , and can be made entirely in terms of real units of goods. Denominating construction costs in final goods simplifies the algebra of the labor market clearing condition. Alternative specifications with costs denominated in labor are possible.

<sup>21</sup>In the continuous space, the market for land is competitive, and marginal revenue equals rent. In a discretized version of the model, landowners in a finite set of points will each accommodate a mass of firms, and face a downward sloping demand curve due to firm heterogeneity. They therefore will act as quasi-monopolists in such a setup.

<sup>22</sup>The second order condition is satisfied by the assumption that  $c''(h) > 0$ .



This allows the landowner's dual decision, choosing both firm density and rents at her location, to be expressed solely in terms of her choice of rent, as higher rents guarantee higher profits conditional on optimal density provision.

But recall that landowner profits holding  $h$  constant are strictly increasing in  $\phi$ , and that this choice is constrained by the participation constraints of firms. Thus, the landowner's constrained optimization chooses highest rents possible, conditional on the willingness to pay of firms. A landowner's location is differentiated from others' by the location-specific parameter  $\eta$ . Each firm's willingness to pay for a space at location  $\eta$  will be conditional on the firm's outside options, including the rent  $\phi'$  at outside options  $\eta' \in S \setminus \eta$ , and variable profits at  $\eta$  and locations  $\eta'$ . A firm of productivity  $\psi$ 's willingness to pay for space at location  $\eta$  can be expressed as

$$WTP(\psi, \eta) = \min_{\eta' \in S \setminus \eta} \{ \psi \cdot (\eta - \eta') + \phi' \}.$$

Intuitively, given a single outside option  $\eta'$ , a firm is willing to pay rent at  $\eta$  equivalent to the difference in variable profit that firm would collect at location  $\eta$  and what would be the variable profit at location  $\eta'$ , paying whatever rent is charged by the landowner at  $\eta'$ . The firm's willingness to pay for  $\eta$  is therefore the minimum difference between profits at  $\eta$  and profits at all other outside options  $\eta' \in S \setminus \eta$ , taking rents at those locations as given.

The landowner will choose the highest possible rent conditional on some firm type being willing to locate at her location. This is equivalent to choosing the firm with the maximum willingness to pay for her location.

$$\begin{aligned} \pi_l &= \max_{\phi} \{ h\phi - c(h) \} \\ \text{s.t. } \phi &\leq \max_{\psi} \{ \min_{\eta' \in S \setminus \eta} \{ \psi \cdot (\eta - \eta') + \phi' \} \}. \end{aligned}$$

Notice that landowners differ only according to their location parameter  $\eta$ . In the equilibrium, the decisions of landowners and firms will jointly determine the matching function  $\psi(\eta)$  between firm-types and the location index.

### 1.2.3 Equilibrium

In this section, I first introduce  $\eta$  as a change of variable and define the rent function  $\phi(\eta)$  and the matching function  $\psi(\eta)$ . I then show that, conditional on an equilibrium existing, sorting always exists. I derive the equilibrium conditions governing the functions  $\psi(\eta)$ ,  $h(\eta)$ , and  $\phi(\eta)$ , and then prove that at least one equilibrium always exists by showing a solution for the mapping function of locations to indices  $\eta(i)$  always exists. An equilibrium will be characterized by a function  $\eta(i)$  that relates market access to locations in  $S$ , a mapping of firms to locations according to locations' market access  $\psi(\eta)$ , a rent curve  $\phi(\eta)$ , and a firm density function  $h(\eta)$ .

To find the functions  $\psi(\eta)$ ,  $h(\eta)$ ,  $\phi(\eta)$ , and  $\eta(i)$ , I derive six equilibrium conditions: (1) the firm spatial equilibrium, (2) the land development equilibrium at each location, (3) the real estate market clearing condition, (4) the labor market clearing condition, (5) the worker spatial equilibrium, and (6) the goods market clearing and balanced trade condition. The analysis herein will first assume an assignment  $\eta(i)$ , solve for  $\phi(\eta)$ ,  $h(\eta)$ , and  $\psi(\eta)$ , and then return to solve for  $\eta(i)$ .

#### Change of variable

I begin by positing a mapping  $\eta(i)$  which characterizes all occupied locations according to  $\eta$ .  $\eta$  serves as index that transforms the space of all occupied locations<sup>23</sup> into a space  $[\underline{\eta}, \bar{\eta}]$ , with boundaries defined by the maximum and minimum values of  $\eta$  in a given equilibrium. For each value of  $\eta \in [\underline{\eta}, \bar{\eta}]$ , a certain density of locations, expressed as  $f(\eta)$ , will share this value.

As an illustration, consider a two-dimensional circular geography with a unit radius, where transportation costs are linear with distance traveled, and all goods must travel through the center of the circle. In this geography, locations closer to the center always have lower transport costs to all other locations and therefore in any equilibrium,  $\eta$  will be

---

<sup>23</sup>Given the assumptions of the model, in particular that  $c(h)$  is continuous and  $c(0) = 0$ , the equilibrium must always be regular and this transformation is always  $S \rightarrow [\underline{\eta}, \bar{\eta}]$ .

decreasing with distance to the center, with all locations on a fixed radius away from the center sharing the same index value  $\eta$ . For any radius  $r$ ,

$$\eta(r) = (1 - r)\bar{\eta} - r\underline{\eta}$$

with a density function  $f(\eta)$  defined as

$$f(\eta) = 2\pi \frac{\bar{\eta} - \eta}{\bar{\eta} + \underline{\eta}}.$$

### Rent gradient and firm sorting

Next, I introduce the rent function  $\phi(\eta)$ , characterizing the rents at each location according to the location index. Because the landowners' optimization problems differ only by  $\eta$ , landowners with identical  $\eta$ 's choose identical rents;  $\eta$  is a sufficient statistic for  $\phi$ .

In order to characterize the rent function, I now introduce the firms' incentive compatibility constraints. In a spatial equilibrium, no entrepreneur can increase her implicit utility by changing locations. Deviations from optimal locations could potentially increase profits in one of two ways: an entrepreneur could move to a location with higher variable profits, or to a location with lower rent. Restating the firm's profit maximization implies neither of these options increase real profits for any entrepreneur.<sup>24</sup> Formally,

$$\forall \psi \in \Psi, \phi(\eta) - \phi(\eta_L) \leq \psi(\eta_L - \eta), \forall \eta_L < \eta \quad (1.2)$$

$$\forall \psi \in \Psi, \phi(\eta) - \phi(\eta_H) \geq \psi(\eta_H - \eta), \forall \eta_H > \eta \quad (1.3)$$

where  $\eta$  is the location chosen by entrepreneur of productivity  $\psi$ ,  $\eta_L$  is an outside option with  $\eta_L \leq \eta^*$ , and  $\eta_H$  is an outside option with  $\eta_H \geq \eta^*$ . Equations (2) and (3) are the

---

<sup>24</sup>The firms' basic tradeoff between variable profit and rent is in terms of baskets of goods so that real profits and implicit utility are equivalent.

incentive compatibility constraints for firms optimally locating in spaces with an index value of  $\eta$ .

It immediately follows from equations (2) and (3) that  $\phi(\eta)$  is strictly increasing. If, instead, a local minimum existed at  $\eta_1 \in (\underline{\eta}, \bar{\eta})$ , all firms at locations to the left of  $\eta_1$  could capture higher variable profit at locations with reduced rent. The firms' incentive compatibility constraints would then guarantee such locations would be unoccupied, contradicting the definition of all locations in  $[\underline{\eta}, \bar{\eta}]$  as the set of occupied locations.

Next, I rewrite the entrepreneur's optimization problem in terms of  $\eta$  and the profit function  $\phi(\eta)$ ,

$$\pi_f = \max_{\eta \in [\underline{\eta}, \bar{\eta}]} \{ \psi \cdot \eta - \phi(\eta) \} ,$$

which yields the first order condition

$$\hat{\psi}(\eta) = \phi'(\eta) \quad \forall \eta \in [\underline{\eta}, \bar{\eta}]. \quad (1.4)$$

Equation (4) defines the mapping of firms  $\psi$  to locations  $\eta$ .<sup>25</sup> Appendix 2 verifies that given a matching function  $\psi(\eta)$  and rent gradient  $\phi'(\eta)$ , landowners at every  $\eta \in [\underline{\eta}, \bar{\eta}]$  choose rents such that equation (4) is satisfied. Because (4) is derived from the firm's optimization, and firm's optimization is the constraint on the landowner's price decision, it exactly satisfies the latter.

Finally, note that for firms to be at their optimum, the second order condition  $\phi''(\eta) > 0$  must hold, which together with equation (4) implies that  $\psi'(\eta) > 0$ . This last result guarantees the positive assortative matching between firms and locations, and as expressed

---

<sup>25</sup>This condition can also be derived from the incentive computability constraints using the Mirrlees conditions. Following Mirrlees (1976), we can further restrict the relevant outside options  $\eta_H, \eta_L$  to be locations with the value of  $\eta$  closest to  $\eta^*$ . Intuitively, if the firms have optimized, deviations from the optimum are increasingly detrimental, and therefore attention can be restricted to local deviations. Thus, for any choice set over which  $\eta$  is continuous, equations (3) and (4) can be evaluated by taking the limit as  $\eta_H \rightarrow \eta_L$ . In the limit, both inequalities bind, becoming the mapping function in equation (4). In this way, the matching between firms and landowners is analogous to a mechanism design problem with a continuum of types with private information.

in the previous sub-section, is a product of the super-modularity of the production function. Proposition 1 formalizes this result

*Proposition 1 : Firm sorting.* In any equilibrium, a strictly increasing function  $\psi(\eta)$  exists; its inverse  $\eta(\psi)$  is weakly positive. That is, in any equilibrium of rents, firm, and worker locations, for any  $\eta_1 > \eta_2$ ,  $\psi(\eta_1) > \psi(\eta_2)$ . Furthermore, in any equilibrium of the above model, a non-degenerate distribution of  $\eta(i)$  must exist for some locations in  $S$ . The only stable equilibria exhibit “strong” sorting, such that the one-to-one function  $\psi(\eta)$  is continuously increasing.

Appendix 1 first proves the assumptions of the model prohibit a degenerate equilibrium distribution of  $\eta(i)$ , then that a location with a higher  $\eta$  must be captured by firms with higher productivity parameter  $\psi$ .

The appendix further elaborates on the possible scope of exceptions to sorting. Any equilibrium must display “weak” sorting properties: the function  $\psi(\eta)$  must be weakly increasing. When the function  $f(\eta)$  is non-singular, sorting is strict. Singularities in  $f(\eta)$  imply a positive mass of locations with a single value of  $\eta$ , and therefore a range of firms at that value of  $\eta$ . However, two firms of the same productivity may not be found at locations with different  $\eta$ ’s. The appendix also shows that any weak sorting equilibrium is unstable.

The remaining analysis assumes a stable one-to-one matching of firms to location productivity. The appendix also notes small changes to the leading predictions of the model in the case of a weak sorting equilibrium, and shows the corollary to Proposition 1, that uniform density is never an equilibrium characteristic, and therefore agglomeration is a pervasive characteristic of this model.

## Density gradient

Next, I rewrite the firm’s profit maximization condition using the rent gradient

$$\tilde{\pi}_l = \max_h \{h \cdot \phi(\eta) - c(h)\}.$$

The first order condition of the above equation implicitly defined the density provision of landowners as a function of  $\eta$  :

$$h(\eta) = c'^{-1}(\phi(\eta)). \quad (1.5)$$

Equation (5) sets the density of firms at a given location  $i$ , given the type of firm at location  $i$ , such that the cost of accommodating the marginal firm is equal to the willingness of that firm to pay for space at  $i$ . While the firm's optimization pins down the rent gradient, the landowner's optimization adjusts the density of firms at each point, thereby distributing the mass of firms in a space according to the index  $\eta$ . The density function  $h(\eta)$  moves the mass of firms from the space  $\psi$ , where the density is defined according to the distribution  $g(\psi)$  to the index's space.

Because of the assumptions on the cost function,  $h'(\eta) > 0$ , and establishment density is increasing in  $\eta$ . In turn, landowners in more productive locations reap larger real profits, both because they attract more productive firms with higher willingness to pay, and because they optimally accommodate a higher density of firms.

### **Real estate market clearing condition**

The previous two equilibrium conditions are derived conditional on a mapping of  $\hat{\psi}$  into  $\eta$ , which moves the distribution of firms by productivity into a distribution of firms on the space of  $\eta$ . The landowners' optimal provision of density will feed, in equilibrium, back into the assignment of firms to locations, as more density at some locations shifts the mass of firms towards those locations. The matching of firms to locations and rents adjust accordingly so as to accommodate all firms which choose to produce.

Each  $\psi$  has associated with it a specific density of firms,  $g(\psi)$ . The density function  $h(\eta)$  describes the equilibrium density of firms at each location according to  $\eta$ . The function  $h$  and  $g$  are defined on different spaces and their distributions will not necessarily resemble one another.

However, the total mass of firms choosing to produce must equal the total mass of firms accommodated equilibrium,  $H(\eta)$ . Furthermore, the total mass of firms with at least a given

productivity  $\psi$  must be the mass of firms with at least  $\psi$  in equilibrium. This intuition yields the following equilibrium constraint

$$\int_{\underline{\eta}}^{\eta_j} h(\eta) f(\eta) d\eta = \int_{\psi(\underline{\eta})}^{\psi(\eta_j)} g(\psi) d\psi$$

at any location  $j$ . Since this condition holds everywhere, differentiating at  $\eta(j)$  we find

$$h(\eta) = \frac{g(\psi(\eta))\psi'(\eta)}{f(\eta)} \quad (1.6)$$

Equation (6) relates the mapping  $\psi(\eta)$  to the firm density function  $h(\eta)$  and the density function  $f(\eta)$ . For any given allocation of firms to locations, the initial density of firms of a specific productivity type must exist somewhere in the actual space.

### Worker spatial equilibrium condition and wages

Workers must be indifferent between the bundle of goods they can consume at each location. For workers to be indifferent, wages at each location  $i$  must exactly offset differences in the price index so workers can attain the same real wage  $\lambda$  across locations.

$$w(i) = P(i) \cdot \lambda$$

Setting wages at a single location  $j_1$  as the numeraire wage, the real wage can be expressed as  $\lambda = \frac{1}{P(j_1)}$ .

### Labor market clearing condition

All labor  $L$  must be used in production. For a given firm type  $\psi$ , the labor bill is equal to the amount of revenue each produces minus variable profits.

$$l(\eta) = \frac{\sigma \cdot p}{\lambda} \cdot \psi(\eta) \cdot \eta$$

Summing over all producing firms and dividing by wages, the total amount of labor must equal the local labor supply,  $L$ .

$$L = \int_{\hat{\psi}}^{\hat{\psi}_u} \left[ \frac{\sigma \cdot \rho}{\lambda} \cdot \psi \cdot \eta(\psi) \cdot g(\psi(\eta)) \right] d\psi(\eta). \quad (1.7)$$

Where  $\underline{\psi}$  is defined by  $\underline{\psi} \equiv \phi(\underline{\eta})/\underline{\eta}$ .

### Solving the mapping $\psi(\eta)$

Equations (4), (5), (6), and (7) jointly determine, for any ordering  $\eta$ , the set of firms that produce, the arrangement of firms into locations, the matching of productivity to locations, the density of firms at each location, and the rent at each location. Putting equilibrium conditions in equations (4)-(7) together, I find the differential equation governing  $\psi(\eta)$ :

$$\psi(\eta) = c'' \left( \frac{g(\psi)\psi'(\eta)}{f(\eta)} \right) \cdot \left[ \frac{f(\eta)g(\psi)\psi''(\eta) + f(\eta)g'(\psi)\psi'(\eta)^2 - g(\psi)\psi'(\eta)f'(\eta)}{f(\eta)^2} \right]. \quad (1.8)$$

Note that the boundaries of the location productivity mapping function,  $\bar{\eta}$ ,  $\underline{\eta}$ , as well as the function  $f(\eta)$  are not yet defined.

In the space of  $\eta$ , the four equilibrium conditions provide the solution to an equilibrium where  $\eta(i)$  is the mapping of locations to  $\eta$ . It remains to be shown that such an equilibrium exists. To do this, I introduce the last two equilibrium conditions, the worker spatial equilibrium the a balanced trade condition, and use them to solve for  $\eta(i)$ .

### Balanced trade

Finally, because all three agents consume at their given or equilibrium locations, trade is balanced, i.e. total market for goods at each location must equal the total amount produced at the location. The local demand at  $j$ ,  $R(j)$ , is therefore equal to the total revenue of all firms at location  $j$ . Recall each firm at location  $i$  receive revenue

$$R(j) = P(j) \cdot r(\psi, j) \cdot h(\eta) = P(j) \cdot \psi \cdot \eta(j) \cdot \sigma \cdot h(\eta) .$$



### Solving for the mapping $\eta(i)$ and the price index.

Equations (4) through (8) characterize any equilibrium in the space of  $\eta$ . While these equations are sufficient to derive predictions. However the general equilibrium requires further mapping of each physical location in  $S$  to a value for the index  $\eta$ . I now use the balanced trade condition to solve for this mapping.

To solve for  $\eta(i)$ , I rewrite the equation for  $\eta(i)$  imposing the balanced trade and worker spatial equilibrium conditions together with the mapping of firms to locations  $\psi(\eta)$ . This yields

$$\eta(i) = P(i)^{-\sigma} \int_{j \in S} \frac{\tau(i, j)^{1-\sigma}}{\rho^{1-\sigma} P(j)^{-\sigma}} \cdot \frac{g(\psi(\eta(j))) \psi'(\eta(j))}{(f(\eta))^2} \cdot (\psi(\eta(j)))^2 \cdot \eta(j) dj \quad (1.9)$$

while the price index can now be expressed as

$$P(i)^{1-\sigma} = \lambda \cdot \int_{j \in S} \left[ \frac{P(j) \tau(i, j)}{\rho \psi(\eta(j))} \right]^{1-\sigma} g(\psi) \cdot \psi'(\eta(j)) \cdot dj \quad (1.10)$$

where  $\lambda = \frac{1}{P(j_1)}$ . Equations (9) and (10) constitute a system of nonlinear Hammerstein equations with a kernel of  $\tau(i, j)^{1-\sigma}$  the solution to which determines the mappings  $\eta(i)$  and  $P(i)$ .

Finally, note that for any equilibrium mapping of  $\eta(i)$ ,  $f(\eta)$  ensures the mass of points in  $S$  is accounted for in  $[\underline{\eta}, \bar{\eta}]$ , such that

$$\int_{\underline{\eta}}^{\bar{\eta}} f(\eta) d\eta = \int_{j \in S} 1 \cdot dj.$$

*Proposition 2* : An equilibrium exists and is described by equations (8)-(10).

*Proof* : See appendix 3.

Appendix 3 proves the system exhibits at least one nontrivial solution and provides conditions for uniqueness. However, because the uniqueness conditions are not easily verifiable, and because they don't on their help provide predictions, I will not assume they are met through the remainder of the paper.

### 1.3 Predictions of the model

In this section, I use the model laid out in the previous section to derive predictions. In doing so, the full advantage of the location index is made plain. As in other geography models,<sup>26</sup> the interdependence of geographic decisions among many agents—in this case firms, workers, and landowners at each location—does not yield a closed-form solution and creates the potential for multiple equilibria. Because of these features, no predictions can be made for the relationship between geographic variables and location or firm characteristics. However, because the index functions as a sufficient statistic both for the location's endogenous characteristics and the characteristics of firms at the location, the model yields relationships between observable location and firm characteristics and the index that will hold in any equilibrium. I use these to derive predictions for key relationships between location characteristics and firm characteristics.

These predictions could equally have been derived from geography-free models with agglomeration forces, and as a result they should not be considered tests of the sorting hypothesis or the geographic mechanisms of the model. Rather, they display the ability of the model, and the indexing method, to derive predictions in a geographic setting without solving the second-stage mapping between  $\eta$ , the location index containing all the geographic information, and locations  $i$ .

To test the sorting hypothesis, I then introduce the composition effects. These effects predict that shocks to density at a given location result in a change in productivity to firms at that location and at neighboring locations based on a change in the sorting pattern of firms.

I isolate two cases in which the direction of the change in productivity runs opposite the direction predicted by the agglomeration forces hypotheses. In particular, average productivity of firms at the most advantageous locations is reduced when those locations experience a positive shock to density. When density is positively shocked anywhere,

---

<sup>26</sup>See, for example, Allen and Arkolakis (2013) or Kyriakopoulou and Xepapadeas (2013).

the average productivity at less-advantageous neighboring locations is reduced, while the productivity at more advantageous neighboring locations increases. (This last effect accords with the agglomeration forces hypotheses). Because the first two predictions can't be derived from models where firms are ex-ante homogenous and density increases cause increases in firm productivity, they constitute a test of the sorting hypothesis.

### 1.3.1 Predictions for firm and location characteristics

In the set of equilibria derived in the previous section, the relationship between specific geographic locations and economic variables remains indeterminate. Underlying geography alone does not determine all subsequent economic decisions. Geographic determinism may be too restrictive a condition for any realistic model. In addition, there is no closed-form solution to the mapping of locations  $i$  to the index  $\eta$ . These features of the model, common in geography models where networked geographic connections affect decisions, inhibits the derivation of predictions on observables.

Despite this, because the location index is a sufficient statistic for both the locations' characteristics and the characteristics of firms locating there, the indexing method yields predictions. Equations (4)-(10), generate relationships between observable location and firm characteristics and the location index. I show how predictions on key relationships between firm and location characteristics can be derived from these equations. These relationships are summarized in the following table

|              |              | Location: | Rent         | Est. den  | Emp. den  | Prices |
|--------------|--------------|-----------|--------------|-----------|-----------|--------|
|              |              | $\eta$    | $\phi(\eta)$ | $h(\eta)$ | $d(\eta)$ | $P$    |
| Firm:        | $\eta$       |           | +            | +         | +         | No     |
| Productivity | $\psi(\eta)$ | +         | +            | +         | +         | No     |
| Size         | $l(\eta)$    | +         | (+)          | +         | (+)       | No     |
| Profits      | $\pi(\eta)$  | +         | (+)          | +         | (+)       | No     |

where  $d(\eta)$  is employment density at locations with index value  $\eta$ , which has not been previously defined. Relationships in parentheses will not be explored in this paper. Note

that there are no relationships between prices and other observable location and firm characteristics. Prices constitute one of the geographic components that form  $\eta$ , and can only be pinned down in a general equilibrium. Because predictions come from only those variables which can be expressed with respect to  $\eta$ , the model yields no predictions with respect to prices and other variables dependent on networked geographic interconnections.

Finally, note that models of agglomeration forces, where density positively affects productivity, predict the following relationships, with the exception of the positive relationship between density and profitability.<sup>27</sup> In particular, with the exception of predictions on firm profits, all the above predictions hold for the model derived in Appendix 5, where I re-derive  $\eta$  using a combination of Marshallian agglomeration forces and pecuniary externalities.

### **The establishment density-productivity relationship**

Using the equilibrium conditions derived in Section 2, I show the model's primary prediction: the productivity-density relationship. Unlike previous theories of firm sorting, this is a prediction, not an assumption of the model. The relationship arises because the most advantageous locations simultaneously attract the most productive firms and command the highest prices, inducing landowners to allow higher establishment density. Both landowners and firms react to the location's characteristic, and the establishment density-productivity relationship arises out of their correlated decisions.

To derive this relationship, I start by introducing the relationship between rents and location advantage and firm quality and location advantage, which will be used here and in the following predictions as well.

*Lemma 3:* In any sorting equilibrium, the mapping  $\hat{\psi}(\eta)$  is increasing  $\partial\hat{\psi}(\eta)/\partial\eta > 0$ , and the rent gradient is convex  $\partial^2\phi(\eta)/\partial\eta^2 > 0$ . Firm density is increasing in  $\eta$ ,  $\partial h(\eta)/\partial\eta > 0$ .

Lemma 3 follows directly from the first and second order conditions of the firm and landowner optimization problems. As ensured by super-modularity, more advantageous

---

<sup>27</sup>In models where firms are ex-ante homogenous, and location effects create productivity differences, spatial equilibrium dictates that firms must pay fixed costs that offset the effect of higher productivity on profits.

locations are matched to more productive firms. From equation (4), it is clear that this alone ensures the rent gradient is increasing and convex in  $\eta$ . Because, by equation (4), the rents are increasing in  $\eta$ , and landowners provide density such that the increasing marginal cost equals the rent, more advantageous locations must have higher density.

Next, I decompose the relationship between firm productivity and density into the relationship between productivity and location advantage, and density and location advantage.

$$\frac{d\psi(\eta)}{dh(\eta)} = \frac{d\psi(\eta)}{d\eta} \cdot \left( \frac{dh(\eta)}{d\eta} \right)^{-1}.$$

*Prediction 1:* Locations with higher firm density have more productive firms:  
 $d\psi(\eta)/dh(\eta) > 0$ .

The model's first prediction immediately follows from the above equation and Lemma 3.

### **The establishment density-employment relationship**

The model predicts a positive relationship between establishment density and the size of firms. Intuitively, higher demand increases firm size through consumer substitution in two ways. First, more productive firms charge lower prices and capture larger shares of demand, irrespective of their location. Second, holding firm productivity constant, higher  $\eta$  locations increase demand through a combination of lower local variable costs, and lower transportation costs to relatively more markets (higher local demand). Because firms sort, more productive firms grow larger due to a Matthew effect: the most productive firms, larger in their own right, purchase locations that push them to grow even more.<sup>28</sup>

Recall that the labor force hired by a firm at  $\eta$  can be expressed as

$$l(\eta) = \frac{\sigma \cdot p}{\lambda} \cdot \psi(\eta) \cdot \eta.$$

Using the method set out in Section 3.1.1, the above equation leads to the second prediction:

---

<sup>28</sup>This in turn implies that part of the well-known productivity-scale relationship is a result of location-specific effects.

*Prediction 2:* Firms in denser locations will be larger.

Prediction 2 follows from the first order condition in the above equation and the sign of the relationships in Lemma 3. Intuitively, higher  $\eta$  locations have both larger firms and higher firm density, and both effects positively affect the relationship between productivity and employment density.

### **The employment density-productivity relationship (population density/productivity relationship)**

In the model, total employment density at a given location is equivalent to total population density at that location, as there is no commuting. Using the labor market clearing condition expressed with respect to  $\eta$  and  $\psi$  the equality

$$\int_{\underline{\eta}}^{\eta} d(\eta) \cdot f(\eta) d\eta = \int_{\underline{\psi}}^{\psi} \left[ \frac{\sigma \cdot \rho}{\lambda} \cdot \psi \cdot \eta(\psi) \cdot g(\psi(\eta)) \right] d\psi$$

must hold for each variable upper boundary  $\psi(\eta)$  and  $\eta$ . Differentiating both sides and substituting equation (6) yields

$$d(\eta) = \frac{\sigma \cdot \rho}{\lambda} \cdot \psi \cdot \eta(\psi) \cdot h(\eta)$$

Prediction 3 follows from the above equation for employment density and Lemma 3.

*Prediction 3:* More productive firms locate in higher employment and higher employment density areas.  $d\psi(\eta)/dd(\eta) > 0$

Intuitively, employment density is simply the firm size at a location times the density of such firms. This relationship would be positive even without the density response built into the model; if all landowners had a constant, inelastic supply of space, making the number of firms at each location constant, the fact that firms in more advantageous locations are larger ensures that firm productivity increases with employment density. Here, the additional density response only increases the size of the employment-location advantage elasticity.

## Productivity and rent

In the model, firms pay for the advantage of higher demand inherent in higher- $\eta$  locations. In geography models with pecuniary externalities and homogenous firms, price differences would arise between locations without firm sorting, but productivity would be constant across locations. The introduction of heterogeneous productivity amplifies price differences and also ensures that more productive firms location in higher priced areas.

The relationship between rents and firm productivity can be expressed as

$$\frac{d\psi(\eta)}{d\phi(\eta)} = \psi(\eta) \cdot \frac{d\psi(\eta)}{d\eta}.$$

Intuitively, firms in locations that are more advantageous are more productive and locations that are more advantageous are more expensive. Prediction 4 immediately follows from the above equation and Lemma 3.

*Prediction 4:* More productive firms pay higher rents:  $d\psi(\eta)/d\phi(\eta) > 0$ .

## The density-profitability relationship

While more productive firms pay higher fixed costs to operate in more productive locations, their overall profits are higher. Intuitively, for a location  $i$  with a given  $\eta$ , more productive firms have higher sales at  $i$ , yet face the same fixed costs (and variable costs). Firms that are more productive than the firm assigned in equilibrium to  $i$  would therefore have higher profits at  $i$ , yet because of the incentive compatibility constraints, their own location must be more profitable for them. Thus even though fixed costs are higher at high- $\eta$  locations, the more productive firms that locate there do so precisely because those higher fixed costs are not out-weighed by the variable profit gains they make, and their overall profits remain higher than their less productive counterparts at less costly locations. Finally, these locations are also denser, as higher prices induce landowners to increase the amount of density they provide.

The model yields the following, final major static prediction:

*Prediction 5:* Firms in denser locations are more profitable.  $d\pi(\eta)/dh(\eta) > 0$ .

*Proof :* First, note that  $\pi'_f(\eta) > 0$ , which is true by taking the first order condition of  $\pi_f$  and substituting equation (4). With Lemma 3, this guarantees the result in *Prediction 5*.

See footnote 24 for an alternative explanation based on results from the mechanism design literature.

This prediction can only hold in models of firm sorting. Specifically, in models with agglomeration forces and homogenous firms, spatial equilibrium holds that price differences, wage differences, and location productivity must offset each other so that all firms are indifferent between locations. Firms pay for increased productivity with higher location-specific costs. If costs did not offset productive amenities, and profits differed across locations, firms in other locations could do better by relocating. In such models, differences in profitability can only be maintained if entrepreneurs receive location-specific consumption amenities that are higher in unprofitable locations so that entrepreneurial utility equalizes across locations when profits do not.

### 1.3.2 Composition Effects

As with the existent sorting models, the predictions of the model thus far have failed to empirically distinguish the sorting hypothesis from models of agglomeration forces, as these alternatives can account for nearly all of the statics predictions in section (3.1). In effect, the observational equivalence result has up until now, continued to hold. In this section, I introduce the composition effect as a test capable of breaking the observational equivalence result and empirically distinguishing between the sorting and agglomeration forces hypotheses.

Under the sorting hypothesis, the productivity of firms at a given location is a function of the mapping  $\psi(\eta)$ , itself a function of the rent and density at all other locations. Exogenous changes to the density of firms at a single location can affect the entire distribution. In the following subsection, I examine two specific instances where small, positive shocks to density generate changes to local productivity through composition effects. In the first case,



the composition of firms in the center of the city is negatively affected by positive local changes in density. Second, I show how a shock to density anywhere generates negative effects on the productivity of firms in less advantageous locations and positive effects in more advantageous locations.

### Composition changes in the urban core

Positive changes to the density in the urban core of the model, defined as the highest  $\eta$  locations, reduces the average productivity of firms at those locations. Intuitively, the urban core, defined as the most advantageous set of locations in a given market, house the most productive firms. Positive changes in the density at the urban core must absorb firms that were formerly priced out of the core and less productive than the least productive firm there. Because the model has no density effects, the addition of more firms has no affect on the productivity of existing firms.

I define the average productivity between some cutoff  $\eta_c$  and the most advantageous location  $\bar{\eta}$

$$\tilde{\psi} = \frac{\int_{\psi(\eta_c)}^{\psi(\bar{\eta})} \psi g(\psi(\eta)) d\eta}{\int_{\psi(\eta_c)}^{\psi(\bar{\eta})} g(\psi(\eta)) d\eta}.$$

Equation (6), the real estate market clearing condition, ensures that the total density of firms between productivity levels  $\psi(\bar{\eta})$  and  $\psi(\eta_c)$  is accounted for in the density of firms present between  $\eta$  and  $\eta_c$ .

$$\int_{\psi(\eta_c)}^{\psi(\bar{\eta})} g(\psi(\eta)) d\psi = \int_{\eta_c}^{\bar{\eta}} h(\eta) f(\eta) d\eta.$$

The real estate market clearing condition drives the composition effect. Because more real estate exists, given a positive shock, within the urban core, more firms must enter. For the condition to hold, the lower bound must move down so that the full mass of firm increases.

To model a shock to the supply density, for some location  $\eta_1 \in [\eta_c, \bar{\eta}]$  I assume an idiosyncratic cost of development at that location, introducing a new parameter  $\kappa \leq 1$  which

affects construction costs for all locations  $\eta \in [\eta_1, \bar{\eta}]$ . Formerly, construction costs everywhere were identically defined according to the function  $c(h(\eta))$ . Now, the construction costs are redefined as

$$c_{new}(h(\eta)) \equiv c(h(\eta)) + \kappa(h) \cdot \epsilon.$$

with  $\epsilon$  arbitrarily small and for a function  $\kappa(\eta)$  defined as

$$\kappa(\eta) = \begin{cases} 0 & \eta < \eta_1 \\ a[h(\eta) - h(\eta_1)] & \eta \geq \eta_1 \end{cases}.$$

for some  $a$ . This functional form hypothesizes an arbitrarily small shock to density while preserving the smoothness of the functional gradients.

The cost of providing density above  $\eta_1$  deviates from the otherwise symmetric cost across the rest of the space. A negative  $a$ , by equation (5) has a positive effect on density at  $i$ . Furthermore, the real estate market clearing condition now becomes

$$\int_{\psi_{new}(\eta_c)}^{\psi(\bar{\eta})} g(\psi(\eta)) d\psi = \int_{\eta_c}^{\bar{\eta}} h_{new}(\eta) d\eta.$$

where  $h_{new}(\eta)$  is the new density function and  $\psi_{new}(\eta_c)$  is the new cutoff firm productivity. In particular  $h_{new}(\eta) > h(\eta)$  when  $a < 0$ .

To accommodate the new density, the left-hand side of the condition must also increase, which is to say the total mass of firms between  $\bar{\eta}$  and  $\eta_c$  must expand. But this can only happen by lowering the lower bound,  $\psi_{new}(\eta_c) < \psi(\eta_c)$ . In order to accommodate more firms in the same space, new, less productive firms that were previously priced out must enter. Prediction 6 follows:

*Prediction 6:* Reductions in building costs in the urban core, defined as the most advantageous locations, by increasing density, decrease average productivity in those locations.  $d\tilde{\psi}/da \leq 0$ .

Note that the prediction holds weakly. As the sorting pattern changes so that less productive firms enter at each location, the quality of firms at those locations decreases marginally,

reducing the density provided to them by the landowners. This attenuates the initial shock, however it cannot reverse the direction.

Note that without the index  $\eta$ , the composition effect only could not give clear predictions. Specifically, the urban core is defined with respect to  $\eta$ . Non-marginal changes in density have the potential to affect advantage of locations in unpredictable ways through geographic general equilibrium effects; increases in density at one location increase the size of local markets for nearby firms and affect price index at all locations. Without closed form solutions, it is impossible to predict the effects of density shocks as mediated through such geographic interconnections. The ability to define the urban core through the location index that can be held constant through marginal changes is crucial for this prediction.

The sign of this prediction obviously contrasts with models where firms are ex-ante identical and density causes increased productivity. In these models, with some exceptions, the productivity of firms at  $\eta_1$  would increase, and the productivity of firms at nearby locations would either remain unchanged, or, through spillovers from firms at  $\eta_1$ , increase. This prediction may not hold in theories positing nonlinear relationship between density and productivity, where, at high levels of density, congestion forces caused decreases in productivity.

The direction of composition changes in the urban core is identified because the only margin for adjustment is the lower bound of firm productivity. The same shock to any other subset of locations  $[\eta_{c1}, \eta_{c2}]$  would have two margins of adjustment, from both higher and lower ends of the productivity distribution. As such, the direction of any change would be ambiguous.

### **Composition changes to competing locations**

Although the model has no clear predictions for the direction of composition changes when development costs are similarly shocked in such subsets of the space  $[\eta_{c1}, \eta_{c2}]$ , the model does have predictions for such a shock's effect on productivity at neighboring subsets  $[\eta_{c0}, \eta_{c1}]$  and  $[\eta_{c2}, \eta_{c3}]$ , where  $\eta_{c0} < \eta_{c1} < \eta_{c2} < \eta_{c3}$ . Such shocks negatively affect the

productivity of firms at less-advantageous neighboring locations  $[\eta_{c0}, \eta_{c1}]$ , and increase productivity at more advantageous locations  $[\eta_{c2}, \eta_{c3}]$ .

Intuitively, the new firms between  $[\eta_{c1}, \eta_{c2}]$  may come from either neighboring location, but will be marginal in either location. All firms in  $[\eta_{c0}, \eta_{c1}]$  are priced out of the more advantageous locations. Those now entering those locations from below will be those on the margin, the most productive firms in  $[\eta_{c0}, \eta_{c1}]$ , who were closest to being willing to pay for  $[\eta_{c1}, \eta_{c2}]$ . The reverse is true for firms entering from above: the firms most easily enticed into the new space available at the less advantageous location must have been the most marginal, and therefore least productive in  $[\eta_{c2}, \eta_{c3}]$ . The removal of marginal firms from the upper and lower margins negatively and positively affect the quality of the average firms in  $[\eta_{c0}, \eta_{c1}]$  and  $[\eta_{c2}, \eta_{c3}]$ , respectively.

Average firm productivity for this neighboring set of spaces is defined as

$$\tilde{\psi}_{NL} = \frac{\int_{\psi(\eta_{c0})}^{\psi(\eta_{c1})} \psi g(\psi(\eta)) d\eta}{\int_{\psi(\eta_{c0})}^{\psi(\eta_{c1})} g(\psi(\eta)) d\eta}.$$

If there is any adjustment on the lower margin of firms in  $[\eta_{c1}, \eta_{c2}]$ ,  $d\psi(\eta_{c1})/da \geq 0$ . But this lower cutoff is also the upper cutoff of the neighboring subset  $[\eta_{c0}, \eta_{c1}]$ . In addition, shift in the matching function between  $\eta_{c0}$  and  $\eta_{c1}$  implies a weakly decreasing lower bound,  $d\psi(\eta_{c0})/d\kappa \geq 0$ , as even after adjustments in density provision in  $[\eta_{c0}, \eta_{c1}]$ , additional firms enter from below to compensate for firms leaving to  $[\eta_{c1}, \eta_{c2}]$ . Together, these two margins of adjustment move the average  $\tilde{\psi}_N$  in the same downward direction.

*Prediction 7:* Reductions in building costs between any locations  $[\eta_{c1}, \eta_{c2}]$  reduces average productivity  $\tilde{\psi}_{NL}$  in the set of less-advantageous neighboring locations  $[\eta_{c0}, \eta_{c1}]$ ,  $d\tilde{\psi}_{NL}/da \leq 0$ . The reverse is true for more advantageous locations,  $d\tilde{\psi}_{NU}/da \geq 0$ .

Because only one or both margins of change may be in effect for the productivity boundaries of  $[\eta_{c1}, \eta_{c2}]$ , the signs of the predictions in Prediction 7 is weak.

In a model where proximity to density increases productivity, the additional density should positively affect productivity both at higher and lower density neighboring locations,

holding the density at those locations constant. The differential impact on higher and lower indexed locations is unique to this model.

## 1.4 Evidence

In the following section, I use establishment-level data on US firms in order to test the predictions of the model. The model's static predictions on the relationships between location and firm characteristics match the data. To my knowledge this is the most extensive documentation of these relationships at the tract level using US data. Again, these predictions are not exclusive to the model in this paper; most of these relationships are present in models of agglomeration forces, where density creates productivity differences. For this reason, the documentation of these relationships do not test the sorting hypothesis,

I then use the panel structure of the Census data to test for sorting among firms that move establishments and firms that expand into new markets. Confirming the literature (Gaubert, 2014), I find evidence for sorting at the city level. In addition, I find evidence for sorting within cities. I also use the panel nature of the data to test for density effects. When controlling for firm and industry effects, I find no evidence of density effects. While these results support the sorting hypothesis, they may be driven by sorting on unobservables unrelated to productivity.

Finally, I test for sorting using the composition effects established in section 3. I isolate shocks to construction expenditures – at the city center and at competing tracts – on the productivity of entrants. I instrument for supply shocks to density using inter-city real estate developer linkages. Positive density supply shocks at the city center and at more advantageous neighboring tracts have negative effects on the productivity of entrants. These results conform to the predictions of the sorting hypothesis and is inconsistent with models of agglomeration forces, where density increases the productivity of local firms.

### 1.4.1 Data

I use restricted access US Census Bureau data on all US establishments between 1992-2007. This includes yearly administrative data on employment and payroll from the Census' Longitudinal Business Database (LBD) and US Economic Census data from all economic Censuses from the years 1992, 1997, 2002, and 2007. In addition, I use IRS establishment-level data to supplement yearly sales data in the LBD and for micro-geographic data at the establishment level, including establishment address, zip code, Census Block and Census Tract. Of all establishments in the LBD in each year – numbering between 6.5 and 7.5 million – I am able to assign tract information to roughly 90%.<sup>29</sup> My sample excludes single-employee establishments. Reported sales in retail, wholesale, and non-tradable service sectors may be reflect differences in local price indexes. To mitigate this issue, I restrict my sample only to tradable sectors. My final sample is composed of between 4.5 and 5 million establishments per year. Column one of Table 1.1 reports summary statistics for this sample.

The majority of my analysis will be at the tract level. I supplement this basic sample with public Decennial Population Census data on tract population, housing, and demographics, and more detailed data on a subset of establishments from Census Surveys, including the Annual Survey of Manufactures in order to confirm the robustness of my results.

Population and firm density are constructed using 2000 Population Census population and SSEL yearly count of active establishments over square miles of land in a tract, respectively. I construct output per worker as a measure of firm productivity. While output per worker is a measure of productivity commensurate with the model, it is realistic to assume the measure in the data is affected by capital levels and worker heterogeneity. To ensure my results are not driven by these forces that are unaccounted for in my model, I use value added per worker and gross margin (value added minus payroll), available for establishments responding to Census of Manufacturing, as well as total factor productivity,

---

<sup>29</sup>Tract information is provided for between 30-60% of observations each year. Using address matching across multiple observations, I am able to assign an additional tract information to an additional 40% of firms. For 20%, I use zip codes to impute tract. A remaining 10% of establishments cannot be traced to specific tracts using the data available. A CES paper details the imputation process.

as calculated by Petrin, Reiter, and White (2012), where possible to confirm results. Columns two and three of Table 1 report summary statistics for these sub-samples, respectively.

Non-residential real estate data is taken from responses to rents at the establishment level from the Censuses of Manufacturers. Although rent does not vary by firm output or employment in the model, I compute rent per worker as an additional metric to ensure plant-size differences are not driving my results.

Finally, I isolate two further sub-samples of establishments: establishment relocations and firm expansions. The former group is identified using a methodology slightly more conservative than that of Lee (2008), isolating establishments that are part of multi-unit firms that shut down in one year and open in a new CBSA at least 50 miles away, in which the firm was not previously present, in the following year. The latter group isolates new establishments of multi-unit firms opened in CBSAs, at least 50 miles away, where there was previously no firm presence. Summary statistics for these two samples are provided in columns four and five of Table 1.1.

#### 1.4.2 Evidence on static relationships

The following sub-section uses Census establishment-level data to test the predictions of the model. The cross-sectional relationships are broadly consistent with those predicted by the model.

Predictions 1-5 posit relationships between establishment level outcomes and tract-level variables. To test these relationships, I estimate the following equation

$$\log(F_{it}) = \alpha_0 + \alpha_1 \cdot \log(T_{tr,t}) + \alpha_2 \cdot X_{c,t} + \alpha_3 \cdot X_{i,t} + \epsilon_{it}$$

where  $F_{it}$  is the establishment-level characteristic for establishment  $i$  measured at time  $t$ ,  $T_{tr,t}$  is the tract level characteristic for tract  $tr$ , measured at time  $t$ ,  $X_{c,t}$  is a vector of city by industry by year fixed effects (used in the second specification only), and  $X_{i,t}$  is a set of establishment variables at time  $t$ , including industry-year fixed effects using the full NAICS

**Table 1.1: Summary statistics**

|  | Observations | Mean      | Std. Dev.  |
|--|--------------|-----------|------------|
| Panel 1: Cross-section of establishments |              |           |            |
| <i>Sales</i>                             | 5,351,354    | 5,530,992 | 1,022,721  |
| <i>Employment</i>                        | 5,351,354    | 23.64     | 129.94     |
| <i>Payroll</i>                           | 5,351,354    | 854,970   | 7,710,470  |
| <i>Value Added</i>                       | 784,364      | 7,080,749 | 75,009,740 |
| Panel 2 :Relocated Establishments        |              |           |            |
| <i>Sales</i>                             | 65,146       | 6,129,400 | 6,126,009  |
| <i>Employment</i>                        | 65,146       | 33.96     | 160.69     |
| <i>Payroll</i>                           | 65,146       | 1,302,700 | 7,549,590  |
| Panel 3: Firm Expansions                 |              |           |            |
| <i>Sales</i>                             | 179,956      | 6,668,400 | 5,041,589  |
| <i>Employment</i>                        | 179,956      | 33.215    | 105.305    |
| <i>Payroll</i>                           | 179,956      | 1,314,547 | 6,434,390  |
| Panel 4: All New Establishments          |              |           |            |
| <i>Sales</i>                             | 1,145,664    | 2,566,148 | 4,356,515  |
| <i>Employment</i>                        | 1,145,664    | 16.99742  | 110.6184   |
| <i>Payroll</i>                           | 1,145,664    | 553,311   | 4,708,828  |

Note. Sample one includes manufacturing or business services establishments responding to Economic Censuses between 1992 and 2007, excluding sole proprietorships, for which geographic data is available or could be imputed from address records. Value added is taken from the subset of firms responding to the Census of Manufactures only. Sample two includes manufacturing or business services establishments responding to Economic Censuses between 1992 and 2007, excluding sole proprietorships and single-unit plants, for which geographic data is available or could be imputed from address records and which are categorized as relocated establishments. Relocations are defined as establishments opened between census years in the same industry (4-digit SIC or NAICS) as an “origin” plant within the same firm that closed the prior year. Relocations within CBSAs or between CBSAs but less than 50 miles apart are excluded. Sample three includes established analogously categorized as firm expansions, defined as new establishments that are part of pre-existing multi-unit firms, opening in a CBSA at least 50 miles away from any other establishments of the same firm. Establishment density is computed as the number of establishments per square mile of land at the tract level. Sample four is the subset of sample one that are new establishments, defined as establishments less than 5 years old at the time of response to the Economic Census.

code,<sup>30</sup> age, and cubic polynomials of the establishment’s latitude and longitude to account for potential spatial auto-correlation in the data.

Tables 1.2 and 1.3 reports relationships for each prediction. City-year fixed effects are added to even-numbered columns. Odd-numbered columns express the basic cross-sectional relationship, comparing tracts both within and across cities, while even columns display the relationship exploiting only within-city variation. All standard errors are clustered at the

<sup>30</sup>Because my sample spans the 2002 SIC / NAICS crossover, I use NAICS-year or SIC-year fixed effects for each code. Replicating the results for a range of years with just SIC or just NAICS codes does not affect the results



**Table 1.2: Productivity vs establishment density**

| VARIABLES                        | Log output per worker |                     | Log VA per worker   |                     | Log TFP              |                       |
|----------------------------------|-----------------------|---------------------|---------------------|---------------------|----------------------|-----------------------|
|                                  | (1)                   | (2)                 | (3)                 | (4)                 | (5)                  | (6)                   |
| <i>Log establishment density</i> | 0.050***<br>(0.002)   | 0.043***<br>(0.003) | 0.026***<br>(0.002) | 0.024***<br>(0.002) | 0.003***<br>(0.0006) | 0.0021***<br>(0.0006) |
| <i>CBSA/Year FEs</i>             | No                    | Yes                 | No                  | Yes                 | No                   | Yes                   |
| Constant                         | 7.46                  | 7.1                 | 4.7                 | -3.89               | 3.46                 | 1.55                  |
| Observations                     | 5,351,354             | 5,351,354           | 784,364             | 784,364             | 569,032              | 569,032               |
| R-squared                        | 0.39                  | 0.49                | 0.29                | 0.52                | 0.47                 | 0.48                  |

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Note. Samples include manufacturing or business services establishments excluding sole proprietorships for which geographic data is available or could be imputed from address records. Sample in columns one and two are all such respondents to Economic Censuses between 1992 and 2007. Sample in columns three and four are manufacturing firms in that year range. VA is value added, reported in the Census of Manufacturers and constructed using output minus value of inputs. Sample in columns five and six are all such respondents to the Annual Survey of Manufacturing between 1993 and 2010. TFP measures are taken from Petrin, White, and Reiter (2011). Establishment density is computed as the number of establishments per square mile of land at the tract level. All regressions are at the establishment-year level and control for age of establishment, cubic polynomials for establishment latitude and longitude, industry-year fixed effects for the establishment's full SIC or NAICS code, and tract-level demographic controls including average age, gender, racial composition, education composition, and income levels. All standard errors clustered at the CBSA level.

CBSA level.

Columns one and two report the elasticity between establishment sales per worker and establishment density. The relationship is consistent both across all locations and using within-city variation in column two, and is within the lower-end of the range in the literature, usually between 3% and 8%.<sup>31</sup> Columns three and four repeat the exercise with value added per worker as a measure of firm productivity. The elasticity is about 50% lower, but fairly constant whether examining between or within city variation. Columns five and six report the relationship using TFP and find a lower figure.

In Table 1.3, Columns seven and eight test the firm size - establishment density relationship. The 5% elasticity is consistent both across all tracts and within cities only. Columns nine and ten find a consistent 20% elasticity between productivity and area commercial rents. Columns eleven and twelve and estimate an elasticity of the employment density-productivity relationship, excluding employment in the observed firm, of about 1%.

<sup>31</sup>See Rosenthal and Strange (2004b) for a complete review of previous findings.

**Table 1.3: Other cross-sectional relationships**

| VARIABLES                        | Log employment      |                     | Log output per worker |                     | Log output per worker |                     | Log gross margin PW |                     |
|----------------------------------|---------------------|---------------------|-----------------------|---------------------|-----------------------|---------------------|---------------------|---------------------|
|                                  | (7)                 | (8)                 | (9)                   | (10)                | (11)                  | (12)                | (13)                | (14)                |
| <i>Log establishment density</i> | 0.051***<br>(0.001) | 0.057***<br>(0.001) |                       |                     |                       |                     | 0.023***<br>(0.002) | 0.023***<br>(0.002) |
| <i>Log avg. rent per worker</i>  |                     |                     | 0.010***<br>(0.005)   | 0.065***<br>(0.005) |                       |                     |                     |                     |
| <i>Log employment density</i>    |                     |                     |                       |                     | 0.011***<br>(0.001)   | 0.008***<br>(0.001) |                     |                     |
| <i>CBSA/Year FEs</i>             | No                  | Yes                 | No                    | Yes                 | No                    | Yes                 | No                  | Yes                 |
| Constant                         | -5.04               | -5.66               | 2.8                   | -3.2                | 9.74                  | 7.1                 | 3.66                | -3.55               |
| Observations                     | 5,351,354           | 5,351,354           | 5,351,354             | 5,351,354           | 5,351,354             | 5,351,354           | 784,364             | 784,364             |
| R-squared                        | 0.27                | 0.4                 | 0.41                  | 0.49                | 0.38                  | 0.49                | 0.28                | 0.51                |

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

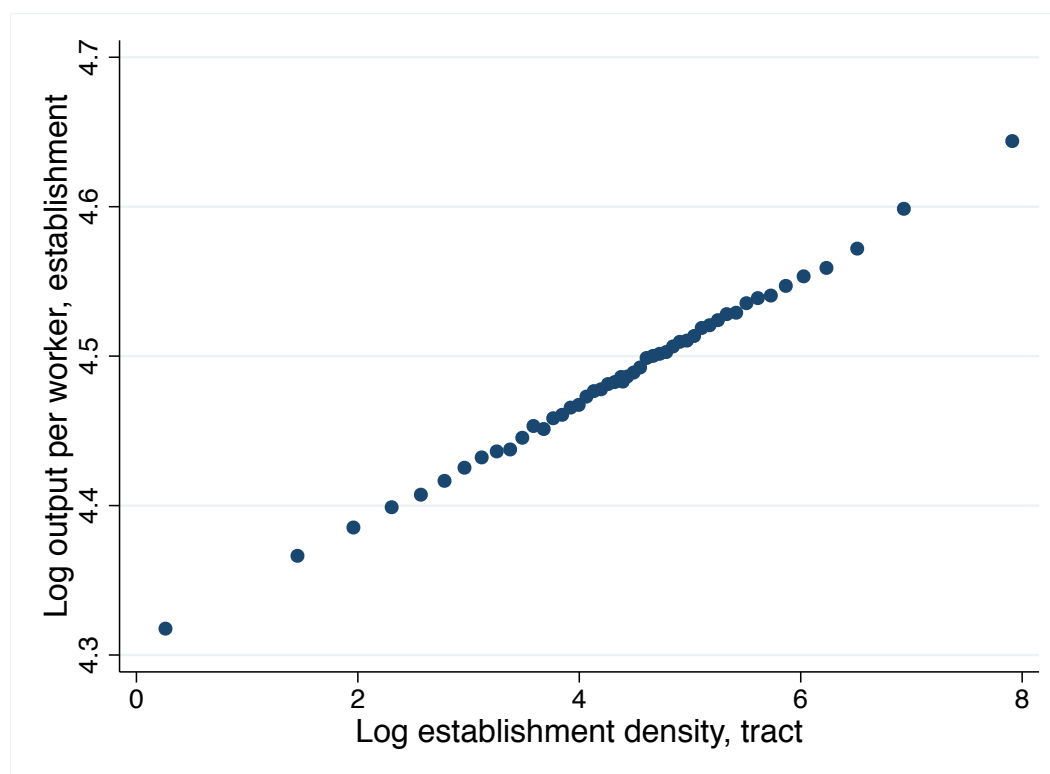
Note. Samples include manufacturing or business services establishments responding to Economic Censuses between 1992 and 2007, excluding sole proprietorships, for which geographic data is available or could be imputed from address records. Establishment density is computed as the number of establishments per square mile of land at the tract level. Average rent per worker is the tract-level average of respondents to questions on rent in the Census of Manufacturers. Employment density is the employees per square mile of land area, excluding those belonging to the observational establishment. Gross margin per worker is value added minus payroll, divided by the number of employees. All regressions are at the establishment-year level and control for age of establishment, cubic polynomials for establishment latitude and longitude, industry-year fixed effects for the establishment's full SIC or NAICS code, and tract-level demographic controls including average age, gender, racial composition, education composition, and income levels. All standard errors clustered at the CBSA level.

Finally, the model predicts that firms in denser locations will receive higher profits. Columns thirteen and fourteen find a positive relationship between log gross margin per employee, measured as value added minus payroll, divided by employment, and establishment density. Gross margin accounts for labor and intermediate inputs but does not account for capital stock. Because the data has poor measures of capital stock at the establishment level, accounting for depreciation without imputation is not possible. If depreciation rates or capital stocks vary significantly by tract density, then this relationship does not capture and is not driven by differences in profitability across density percentiles. This last measure must be taken purely as suggestive of the possibility that profits do indeed vary along with establishment size and worker productivity, across density percentiles.

The positive relationship between gross margin and density has an important secondary implication. Because the model and thus these empirics do not account for differences in labor force quality (such as education attainment), I risk mistaking worker productivity for firm productivity (see Combes *et al.* (2008) for a discussion thereof). If it was the case that worker productivity was driving the empirical relationships explored here, one would

predict a negative or zero elasticity between margins, calculated as value added minus payroll, and density.

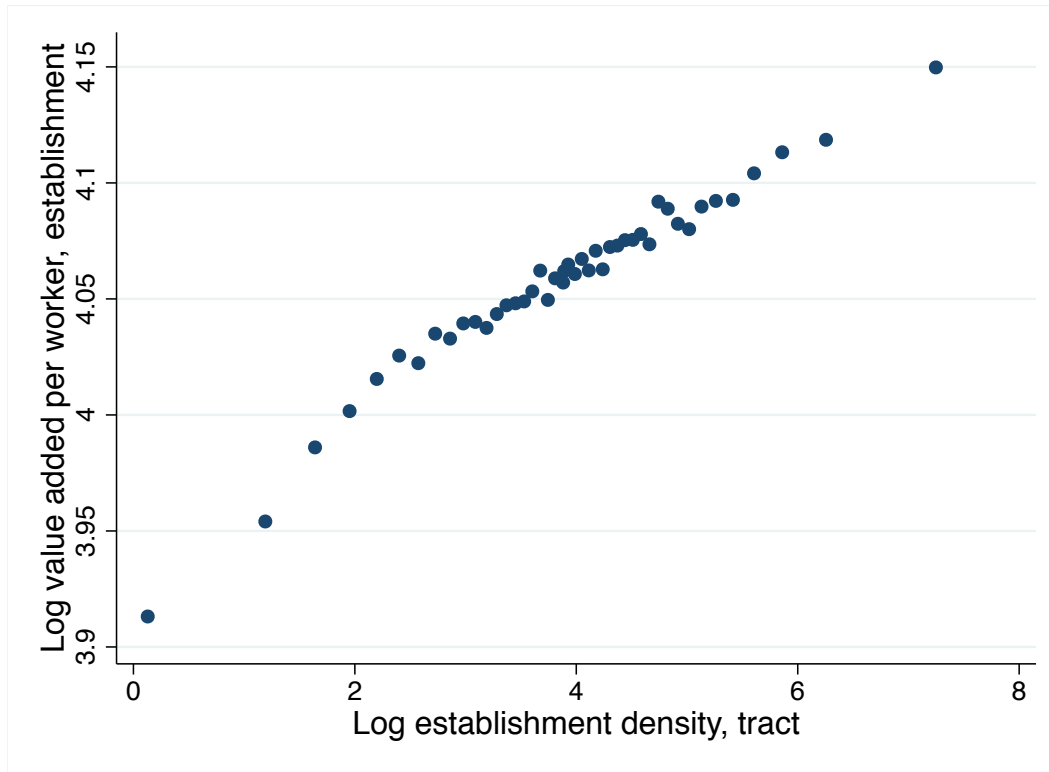
Figures 1.3-1.8 report binned scatterplots where residuals of location and establishment characteristics from these regressions.<sup>32</sup> The scatterplots show that these relationship are remarkably consistent across all density percentiles. Appendix figures replicate these non-linear plots using progressively fewer controls. Appendix Table 2 replicates Table 2 using the entire population of US establishments and finds similar results.



**Figure 1.3:** *Log output per worker vs log establishment density*

---

<sup>32</sup>Note. All figures are binned scattered residuals of tract level variables on establishment level variables, controlling for establishment age, industry- and CBSA-year fixed effects, as well as cubic polynomials in establishment latitude and longitude

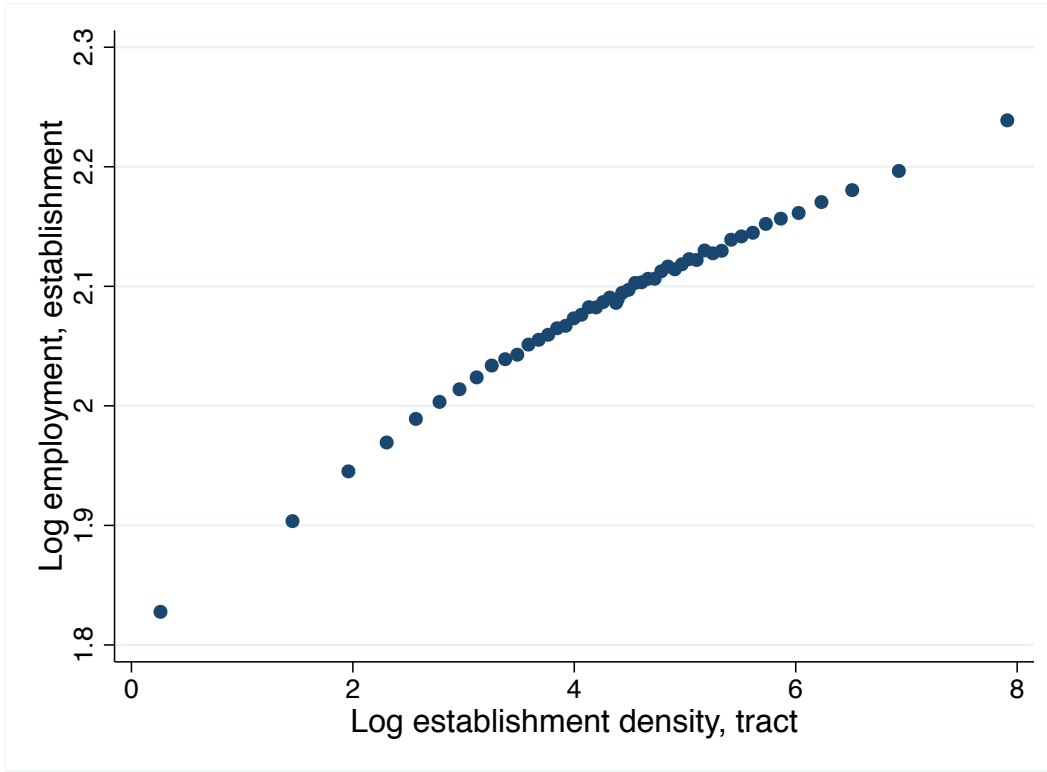


**Figure 1.4:** *Log value added per worker vs log establishment density*

### 1.4.3 Evidence from movers

I use the above model to form predictions on the relocation decisions of plants and expansion decisions of multi-unit firms. The model does not directly make predictions about relocations or expansions: it is static and has no multi-unit plants. Nevertheless, observation of firm characteristics in one period allows me to measure such characteristics when the firm's establishments are divorced from their subsequent locations, making relocations and expansions useful descriptive evidence. I find evidence consistent with the sorting hypothesis, mixed evidence for city-size effects and no evidence for density effects. Above all, it should be stressed that these results do not show causal relationships, as they may be driven by selection on omitted variables unrelated to productivity.

I consider two distinct groups of movers: (1) multi-unit establishments that move at least one plant between cities at least fifty miles away, and (2) multiunit establishments that



**Figure 1.5:** *Log establishment size vs log establishment density*

open new establishments in new cities at least fifty miles away. The data does not allow for the identification of relocations of single-unit firms or expansions of previously single-unit firms.

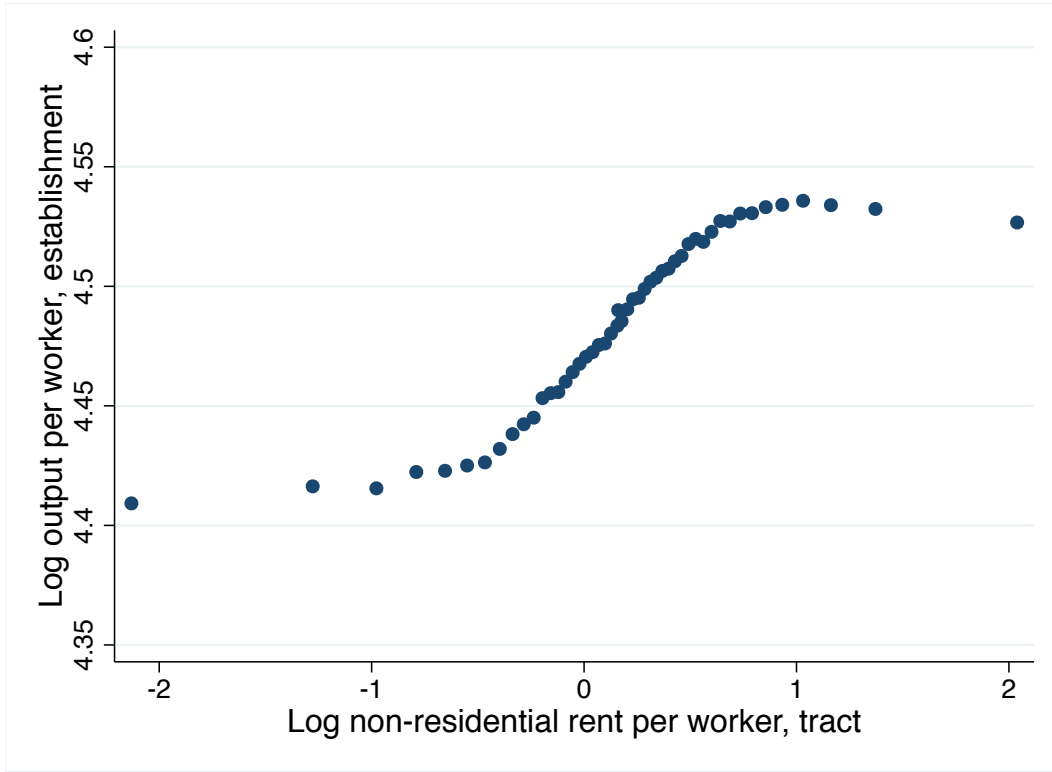
First, I consider multi-unit firms that relocate the production of a single establishment. I follow the literature (Lee, 2008) in identifying plant moves. I populate my sample by identifying only new plants with an identical 4-digit modal industry code as one or more plants within the same firm that closed in a separate city and state in the previous year.

<sup>33</sup>This sample consists of roughly 0.01% of the total population of firms in each year. I estimate

$$\log(h_{tr,e,t}) = \alpha_0 + \alpha_1 \log(\psi_{i,t-1}) + \alpha_2 \cdot X_{e,t} + \alpha_3 X_{f,t-1} + \alpha_4 dist_{e,f} + \epsilon_{e,t}$$

---

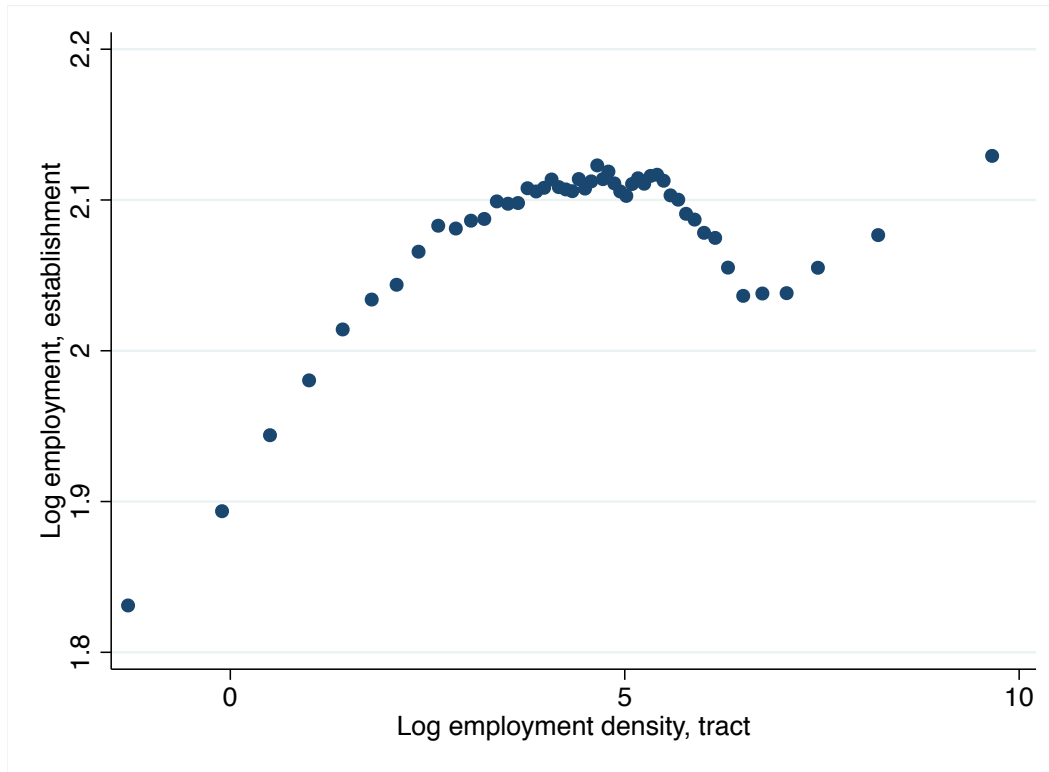
<sup>33</sup>This is a more conservative definition than Lee 2008, who includes new plants within the same industry as existing plants that reduced output by at least 50% in the previous period. For this reason my sample is smaller than his.



**Figure 1.6:** *Log output per worker vs log rent per worker*

where  $h_{tr,e,t}$  is the establishment density of the tract where the relocated establishment  $e$ , as a part of firm  $f$ , exists in period  $t$ ,  $\psi_{f,t-1}$  is the firm's average productivity, calculated as output per worker at each establishment and weighted by establishment size, in period  $t - 1$ ,  $X_{e,t-1}$  is a vector of establishment-specific variables including latitude and longitude cubes, industry-year fixed effects, and in some specifications will include industry by CBSA by year fixed effects,  $X_{f,t-1}$  is a set of latitude and longitude cubes for firm centroid at  $t - 1$ , and  $dist_{e,f}$  is the distance between the firm centroid at  $t - 1$  and the establishment location at  $t$ .

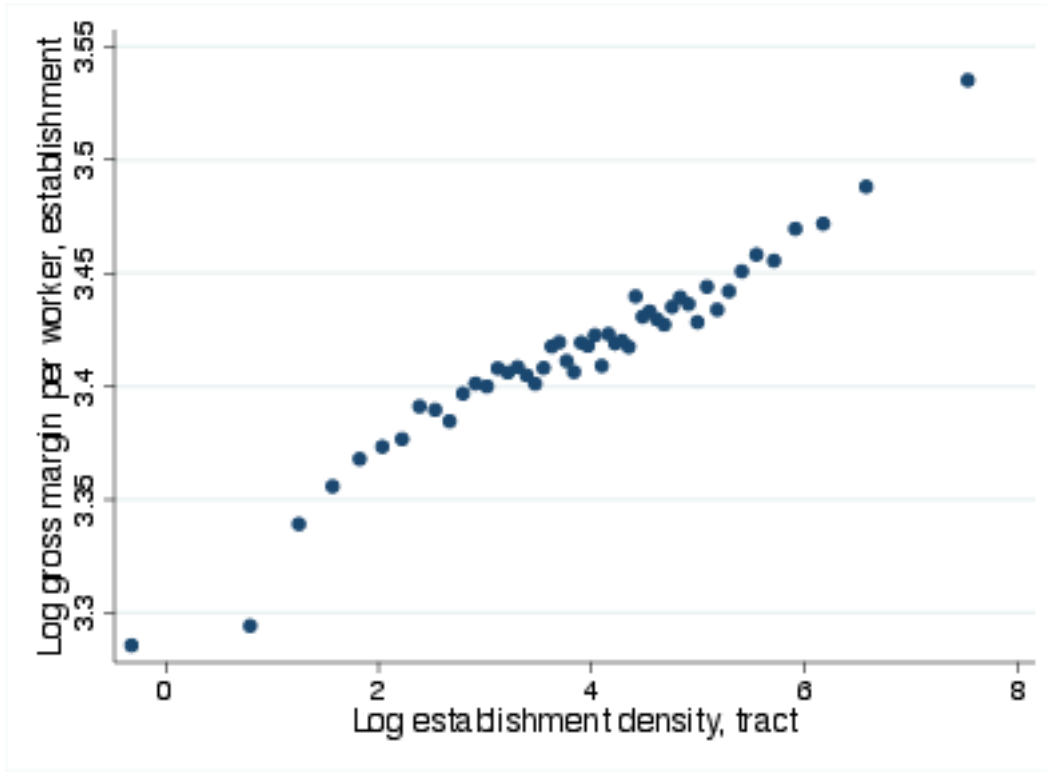
Table 1.4 reports the results. Columns one and two report establishment movers while three and four repeat the exercise for firm expansions. As expected under sorting, output per worker in prior years positively predicts the density at the new location. When CBSA-year fixed effects are taken into account in even columns, the elasticity drops by between half to two thirds, although remains significant for plant relocations.



**Figure 1.7:** *Log output per worker vs log population density*

If firms did not sort on density, a zero or negative result would be expected driven by mean reversion. While suggestive of sorting, these results are limited by two identification problems. First, both sorting and agglomeration forces could be affecting my coefficients, and therefore their magnitude may not reflect the work of sorting alone. Second, the sorting hypothesis proposes sorting based on productivity. Instead, the observed sorting may be based on other firm unobservables that correlate with density. Because both the previous and current period measures of output could be the result of density-driven agglomeration forces, these measures may confound agglomeration forces for sorting on productivity where none exists. For that reason, they must be interpreted with caution.

Table 1.5 directly tests for agglomeration forces by examining how density at new locations affects productivity when conditioning on the firm's productivity in the previous period. I estimate



**Figure 1.8:** *Log gross margin per worker vs log establishment*

$$\log(\psi_{e,t}) = \alpha_0 + \alpha_1 \log(h_{tract,e,t}) + \alpha_2 \log(\psi_{f,t-1}) + \alpha_3 \cdot X_{e,t} + \alpha_4 X_{f,t-1} + \alpha_5 dist_{e,f} + \epsilon_{e,t}$$

where  $\psi_{e,t}$  is the output per worker of the new establishment at time  $t$ , and other variables are as before.

Columns one and four report a positive effect of density on establishment productivity. When city-year fixed effects are introduced in columns two and five, the point estimates fall and the effect becomes zero, suggesting city-size, but not density effects, exist.

The last four regressions test use the multi-plant nature of these firms to control for firm-level productivity-year effects. Columns three and six introduce firm-industry-year fixed effects. Here, I am comparing two relocations within the same firm-industry category in the same year. Columns four and eight add city-year fixed effects. I can reject the initial point estimates. Taken together, these regressions give weak evidence for city-size effects and no evidence for density effects.



**Table 1.4: Sorting on Productivity**

| VARIABLES                               | Log establishment density        |                  |                        |                     |
|---|----------------------------------|------------------|------------------------|---------------------|
|   | (1)<br>Establishment relocations | (2)              | (3)<br>Firm expansions | (4)                 |
| <i>Prior-year log output per worker</i> | 0.028***<br>(0.007)              | 0.004<br>(0.005) | 0.086***<br>(0.014)    | 0.047***<br>(0.011) |
| <i>CBSA FEs</i>                         | No                               | Yes              | No                     | Yes                 |
| Constant                                | 22.44                            | 86.97            | 15.55                  | 64.28               |
| Observations                            | 65,146                           | 65,146           | 179,956                | 179,956             |
| R-squared                               | 0.58                             | 0.68             | 0.6                    | 0.7                 |

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Note. Samples include manufacturing or business services establishments responding to Economic Censuses between 1992 and 2007, excluding sole proprietorships and single-unit plants, for which geographic data is available or could be imputed from address records. Sample in columns one and two use establishment relocations, defined as establishments opened between census years in the same industry (4-digit SIC or NAICS) as an “origin” plant within the same firm that closed the prior year. Relocations within CBSAs or between CBSAs but less than 50 miles apart are excluded. Sample in columns three and four use firm expansions, defined as new establishments that are part of pre-existing multi-unit firms, opening in a CBSA at least 50 miles away from any other establishments of the same firm. Establishment density is computed as the number of establishments per square mile of land at the tract level. All regressions are at the establishment-year level and control for age of establishment, industry-year fixed effects for the establishment’s full SIC or NAICS code, cubic polynomials for both the establishment latitude and longitude and the latitude and longitude of the closed plant or firm’s geographic center, respectively, and the distance between the old and new plants or the new establishment and firm’s geographic center, respectively. All standard errors clustered at the CBSA level.

It should be noted that these specifications do not account for differences in trends; firms which are increasingly productive may also sort into increasingly dense locations. Such effects would bias results positively. Again, these results must only be understood as suggestive.

Online appendix Table 3 and Table 4 repeat the above exercise for all multi-unit movers across all industries. Though they find slightly more evidence for city-size effects and some evidence for density effects, they may be picking up price effects in non-tradable industries.

**Table 1.5: Marginal effects of density**

| VARIABLES                             | Log output per worker, 5 years forward |                       |                    |                     |                       |                     |                    |                     |
|---------------------------------------|--|-----------------------|--------------------|---------------------|-----------------------|---------------------|--------------------|---------------------|
|                                       | (1)                                    | (2)                   | (3)                | (4)                 | (5)                   | (6)                 | (7)                | (8)                 |
|                                       | Establishment Movers                   |                       |                    |                     | Firm Expansions       |                     |                    |                     |
| <i>Log est. density, destination</i>  | 0.0070**<br>(0.0031)                   | 0.0016<br>(0.0033)    | 0.0035<br>(0.0032) | -0.0004<br>(0.0036) | 0.0046**<br>(0.0024)  | -0.0030<br>(0.0025) | 0.0008<br>(0.0020) | -0.0027<br>(0.0024) |
| Firm-level log output pw              | 0.2806***<br>(0.0104)                  | 0.2735***<br>(0.0106) | .                  | .                   | 0.1407***<br>(0.0131) | 0.290***<br>(0.012) | .                  | .                   |
| <i>CBSA fixed effects</i>             | No                                     | Yes                   | No                 | Yes                 | No                    | Yes                 | No                 | Yes                 |
| <i>Origin lat/lon cubes</i>           | Yes                                    | Yes                   | No                 | No                  | Yes                   | Yes                 | No                 | No                  |
| <i>Industry, year fixed effects</i>   | Yes                                    | Yes                   | No                 | No                  | Yes                   | Yes                 | No                 | No                  |
| <i>Firm, industry, year fixed eff</i> | No                                     | No                    | Yes                | Yes                 | No                    | No                  | Yes                | Yes                 |
| Constant                              | 0.6                                    | 7.52                  | 14.5               | 30                  | 2.98                  | 6.84                | 5.5                | -1.3                |
| Observations                          | 40,064                                 | 40,064                | 40,064             | 40,064              | 129,120               | 129,120             | 129,120            | 129,120             |
| R-squared                             | 0.93                                   | 0.94                  | 0.94               | 0.95                | 0.93                  | 0.93                | 0.95               | 0.96                |

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Note. Samples include manufacturing or business services establishments responding to Economic Censuses between 1992 and 2007, excluding sole proprietorships and single-unit plants, for which geographic data is available or could be imputed from address records. Sample in columns through four use establishment relocations, defined as establishments opened between census years in the same industry (4-digit SIC or NAICS) as an "origin" plant within the same firm that closed the prior year. Relocations within CBSAs or between CBSAs but less than 50 miles apart are excluded. Sample in columns five through eight use firm expansions, defined as new establishments that are part of pre-existing multi-unit firms, opening in a CBSA at least 50 miles away from any other establishments of the same firm. Establishment density is computed as the number of establishments per square mile of land at the tract level the year of observed relocation or expansion. Prior census year log output per worker is the average output per worker of all other plants in the year of relocation or expansion. Log output per worker, five years forward is the productivity measure of the new establishment taken five years after the relocation or expansion. All regressions are at the establishment-year level and control for age of establishment, industry-year fixed effects for the establishment's full SIC or NAICS code, cubic polynomials for the establishment latitude and longitude, dummies for the closure of the new plant 5 years after moving, and the distance between the old and new plants or the new establishment and firm's geographic center, respectively. All standard errors clustered at the CBSA level.

## 1.4.4 Composition Effect

Exogenous variation to the supply of density affects the quality of entrants, both in the urban core and at neighboring tracts. In the following subsection, I first propose a strategy for identifying exogenous variation to the supply of density, then test both composition effects. The OLS and IV results support the predictions of the model.

## Instrumental Variable Approach

In order to isolate composition effects effect, I must first isolate exogenous, marginal changes to local construction costs. To do this, I use the inter-city linkages of real estate developers. Commercial real estate development is a leveraged industry (Gyourko, 2009). Developers such as real estate investment trusts, or REITs that also hold an lease real assets may be exposed to real estate shocks via effects on income from these assets. Changes in local real estate prices may therefore affect firms' propensity to supply space.

If such leaser-developers are active in more than one market, local shocks in one market can affect their ability to supply space in other markets. A firm with projects in Boston and Philadelphia, for example, may delay, sell, or scale-down a Boston project due to a negative shock to the market in Philadelphia that depletes the firms' assets. On the other hand, positive shocks to prices in Philadelphia may divert scarce resources away from Boston projects. I term the former effect an income effect and the latter a price effect.

Using the Census of Finance and Insurance, to which all lessors of real estate respond, I isolate over two hundred firms that operate in my period of observation in multiple cities and undertake new commercial real estate construction in at least one. I predict construction expenditures of a given firm's establishment using the construction expenditures and sales of single-unit leaser-developers operating in linked cities, that is, other cities where the developers have an established presence, weighted by their previous-period payroll in each city. Both the income and price effects appear to be operational.

The first four columns of Table 1.6 show the relationship between predicted and actual construction expenditures for tract-level aggregates. Controlling for tract fixed effects in columns, the two are positively correlated and significant. Column two adds city and tract level controls. Columns three and four repeat the exercise using changes.

I use these predicted values as instruments for the level of construction expenditures at the tract and city levels in the following subsection in order to find the effects of these construction expenditures on the quality of entrants. The identifying assumption I make is that the level of sales and investment of single-unit establishments in linked cities affects the quality of entrants in another city only via construction expenditures of the linked firms. Several possible channels may violate this exclusions restriction and bias the results.

First, the exclusion restriction may be violated if linked cities are exposed to correlated real estate shocks. The model itself suggests the decision to enter each market is endogenous to a firm characteristics. If these endogenous links are formed between similar cities, and in particular if linked cities share characteristics that expose them to common price shocks, positive shocks in one market will appear as positive space supply shocks to the linked

**Table 1.6:** *Predicted construction expenditures*

| VARIABLES   | Log construction expenditures, tract level |                     |                    |                     |
|---|--|---------------------|--------------------|---------------------|
|   | (1)  | (2)                 | (3)                | (4)                 |
|   | Levels                                     |                     | Changes            |                     |
| <i>Log predicted sales, tract level</i>           | 0.105***<br>(0.034)                        | 0.104***<br>(0.035) |                    |                     |
| <i>Change in log predicted sales, tract level</i> |  |                     | 0.086**<br>(0.026) | 0.084***<br>(0.026) |
| <i>CBSA size</i>                                  | No   | Yes                 | No                 | Yes                 |
| <i>CBSA/Year FE</i>                               | Yes  | Yes                 | No                 | No                  |
| Constant  | -0.21                                      | -420                | -0.08              | -2.58               |
| Observations                                      | 4,088                                      | 4,088               | 4,088              | 4,088               |
| R-squared   | 0.22                                       | 0.22                | 0.01               | 0.02                |

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Note. Sample includes all tracts for which Real Estate Investment Trusts respond to questions in the Census of Finance and Insurance respondents on expenditures on new construction on commercial real estate, which have establishments reporting sales at least one other CBSA, and for which geographic information was present or could be imputed from address files. Log construction expenditures at the tract level is the log of the sum of the dollar value of all construction expenditures on commercial real estate, including office and manufacturing space, reported. Columns one and two report this measure in levels and columns three and four report changes between census years at the tract level. Log predicted sales are the predictions based on current-year sales of single-unit firms responding to the same census questions in CBSAs where the firm responding in the observed tract has other establishments present, weighted by the percent of sales at the firm level in that CBSA in the previous year. Changes report log differences for this predicted value. All regressions are at the tract level and include controls for year fixed effects. All standard errors clustered at the CBSA level.

market. This will positively bias the sales instrument and negatively bias the expenditure instrument. In addition to correlated shocks, developers may specialize in similar types of markets, especially markets with particular similar trends. Similar trends in each market may appear and act as correlated shocks, biasing the instrument as above. Alternatively, developers may enter markets endogenously to hedge against idiosyncratic shocks. If endogenous links formed for hedging will negatively bias the sales instrument.

Columns five and six of Table 1.6 test for endogenous linkages by estimating the relationship between the tract-level aggregate predicted values for linked developers and city-level construction expenditures of non-linked developers. Appendix figures 1a and 1b show a binned scattered residual plot of these relationships. Relationships are positive but

insignificant. Importantly, there is no evidence that developer linkages are formed to hedge idiosyncratic city-level shocks.

A third and fourth potential for bias enter as development may respond to income and price effects through other margins. The instrument relies on adjustments on the scope and timing of projects. Leaser-developers may select instead to adjust their selection of projects, choosing to delay the lowest-margin projects, or to alter the quality rather than scope or timing of a project. Lower quality sites or sites with lower-quality construction may attract lower quality firms or produce relatively less productive amenities. The former channel negatively biases results while the latter may be a positive bias.

Of these four channels, only site selection and market hedging negatively biases the sales instrument. Because there appears to be no evidence for market hedging, I only use the sales instrument.

### **Shock to supply of space in the urban core**

Shocks to the supply of space in the urban core reduce the average productivity by accommodating the entrance of lower-quality firms.

In order to test this composition effect at the urban core using my sample, several ancillary assumptions must be made. First, for statistical power, I treat each of the 361 CBSAs reporting construction expenditures on commercial real estate as a distinct market, i.e., that no inter-city location decisions are made by firms. As stated in Section 3.2.1, the direction of the composition effect is only certain when the most-productive firm's location is fixed. If firms choose establishment locations within and between cities, more space in the city center may attract more productive firms from other, larger cities, attenuating any negative result.

Proposition 7 implies that the productivity of entrants relative to incumbents is decreasing in a shock to space provided in the urban core. Having isolated each city, the next step must be to identify the urban core, or the most advantageous locations in the city. Because location advantage,  $\eta$ , is a sufficient statistic for all economic activity, the model provides a

natural proxy for  $\eta$ : establishment density. I rank each tract according to the establishments per square mile in each.

Next, the model is agnostic as to the cutoff threshold for the urban core. This provides a natural secondary test: a negative result must be robust to any arbitrary lower threshold. I break the city into density percentiles and test for an negative result in each. Specifically, I estimate

$$\log(\psi_{p,e,t}) = \alpha_0 + \alpha_1 \log(c_{p,t-1}) + \alpha_2 \log(\psi_{p,i,t-1}) + \alpha_3 X_{e,t} + \alpha_4 X_{tract} + \epsilon_{pe,t}$$

where  $\psi_{p,e,t}$  is the output per worker of entrants at time  $t$  within the percentile cutoff threshold  $p$ ,  $\psi_{p,i,t-1}$  is the average productivity of incumbents within the cutoff at time  $t - 1$ ,  $c_{p,t-1}$  is the dollar value of construction expenditures in the previous-period,  $X_{e,t}$  is a vector of establishment controls, including, age, latitude and longitudinal cubes, and industry, and  $X_{tract,t}$  is a vector of tract fixed effects, and where  $p$  is, alternately, the 25th, 50th, 75th, 90th, and 95th percentile. While tract fixed effects are important for removing the baseline productivity differences between tracts, trends in productivity differences are removed by controlling directly for the lagged productivity of incumbents. Proposition 7 predicts  $\alpha_1 < 0$ , contrary to models where density causally improves productivity.

Because construction may respond to changes in demand, I instrument for supply shocks to construction using the predicted expenditures as described in Section 4.4.1, where the first stage estimates

$$\log(\hat{c}_{p,t-1}) = \beta_0 + \beta_1 \cdot \log(pred_{p,t-1}) + \epsilon_{p,t-1}.$$

Table 1.7 reports the results. Columns one, three, five, seven, and nine report the OLS results for the 25th, 50th, 75th, 90th, and 95th percentiles respectively. Coefficients are small and negative, and except for at the 75 percentile cutoff, significant.

Columns two, four, six, eight, and ten report IV results. When instrumenting for changes in supply, the effect of construction is negative, however not always significant. Although a positive effect cannot be excluded at every percentile threshold, the persistent, negative effects are consistent with the sorting hypothesis.

**Table 1.7: Composition effects at urban core**

| VARIABLES                    | Log output per worker |                   |                    |                   |                   |                     |                     |                  |                     |                   |
|------------------------------|-----------------------|-------------------|--------------------|-------------------|-------------------|---------------------|---------------------|------------------|---------------------|-------------------|
|                              | 25th percentile       |                   | 50th percentile    |                   | 75th percentile   |                     | 90th percentile     |                  | 95th percentile     |                   |
|                              | OLS                   | IV                | OLS                | IV                | OLS               | IV                  | OLS                 | IV               | OLS                 | IV                |
|                              | (1)                   | (2)               | (3)                | (4)               | (5)               | (6)                 | (7)                 | (8)              | (9)                 | (10)              |
| <i>Log const. exp.</i>       | -0.007**<br>(0.003)   | -0.004<br>(0.004) | -0.006*<br>(0.003) | -0.003<br>(0.004) | -0.001<br>(0.001) | -0.029**<br>(0.014) | -0.002**<br>(0.011) | 0.011<br>(0.012) | -0.005**<br>(0.002) | -0.007<br>(0.015) |
| <i>Angrist-Pischke value</i> |                       | 6.56              |                    | 8.56              |                   | 16.38               |                     | 14.92            |                     | 21.42             |
| Observations                 | 1,145,664             | 1,145,664         | 883,331            | 883,331           | 452,092           | 452,092             | 150,050             | 150,050          | 61,401              | 61,401            |
| R-squared                    | 0.34                  | 0.38              | 0.35               | 0.37              | 0.33              | 0.38                | 0.46                | 0.38             | 0.39                | 0.35              |

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Note. Samples include all new entrants, defined as establishments less than 5 years old at the time of the Economic Census, in manufacturing or business services industries responding to Economic Censuses between 1992 and 2007 for which geographic data is available or could be imputed from address records, in the listed density percentile or above. Density percentiles are constructed by ranking tracts in each CBSA according to the establishments per square mile of land in the tract. Log construction expenditures is the tract-level sum of reported construction expenditures on new office or manufacturing space, in thousands of dollars, of REITs in the tract, or the instrumented value (see text for IV strategy and procedure). All regressions control for CBSA fixed effects, year fixed effects, age latitude longitude cubic functions, and full industry code fixed effect of establishments, prior-period average log output per worker of incumbent establishments. All standard errors are clustered at the CBSA level.

Although agglomeration forces cannot, on their own, explain the negative result at the 95th and 50th percentiles, this result does not refute their existence. Such forces, acting in conjunction with sorting, may bias the result in either direction. If agglomeration work directly through density, added space would amplify agglomeration forces. Alternatively, if agglomeration forces work through average productivity, the decreased productivity of entrants may amplify a negative sorting effect. A negative result therefore cannot act as a rebuttal of agglomeration forces but only as evidence of the existence of sorting forces.

Additionally, the test here implicitly assumes that agglomeration effects would take place within the first five years of a shock to density. If agglomeration forces take time to develop, perhaps through changes in the structure of the labor market as workers slowly adjust to the shock, the initial negative effects may fade over time and be replaced with the positive effects of agglomeration forces.

At their highest, the IV results predict about a three percent decrease in productivity of entrants for every doubling of construction activity. These results are not enormous but do suggest extremely powerful sorting forces. Congruent with the literature (Gyourko, 2009), I likely observe on average 20% of the newly-built office and manufacturing space in urban

centers in this time period. While my time period reflects a period marked by booming commercial real estate construction, new construction adds to roughly 15%-20% vacancy rates in urban office space across this period. In addition, while most new entrants may be unaffected by changes to supply, the model predicts changes to productivity will occur through the productivity of marginal entrants. It is therefore possible that the observed coefficients are the result of the interaction of sorting forces with agglomeration forces.

Finally, if density and productivity are co-determined by unobserved variables, this result may be attained without sorting. For example, if local public goods improve productivity of firms and increase density while crowding reduces productivity of all firms, more crowding would decrease the productivity of entrants in the absence of both sorting and agglomeration forces. While this specification cannot rule this possibility out, the following specification does.

### **Shock to supply of space in competing tracts**

Shocks to landowner's ability to provide space in any particular location affects the productivity of entrants at neighboring locations differentially based on the relative quality of the shocked location. According to Proposition 10, a shock to density in one location systematically draws the least productive firms from more advantageous locations, improving average productivity at such locations, and the most productive firms from less advantageous locations reducing average productivity at such locations. To test Proposition 8, I use the ordering of tracts by density to estimate the following tract-level regressions

$$\log(\tilde{\psi}_{NL,t}) = \alpha_0 + \alpha_1 \cdot \log(\text{cont}_{tract,t-1}) + \alpha_2 \cdot \log(\tilde{\psi}_{NL,t-1}) + \alpha_3 X_{tract} + \alpha_4 \cdot X_t + \epsilon_{tract,t}$$

$$\log(\tilde{\psi}_{NU,t}) = \alpha_0 + \alpha_1 \cdot \log(\text{cont}_{tract,t-1}) + \alpha_2 \cdot \log(\tilde{\psi}_{NU,t-1}) + \alpha_3 X_{tract} + \alpha_4 \cdot X_t + \epsilon_{tract,t}$$

where  $\tilde{\psi}_{NL,t}$  and  $\tilde{\psi}_{NU,t}$  are the average productivity of entrants to the next-lowest and next-highest density tracts, respectively,  $\tilde{\psi}_{NL,t-1}$  and  $\tilde{\psi}_{NU,t-1}$  are the productivity of incumbents of the tracts at time  $t - 1$ ,  $X_{tract}$  is a vector of tract-level variables including (and limited to at first) tract fixed effects and  $X_t$  are year fixed-effects.



**Table 1.8: Composition effects at neighboring tracts**

| VARIABLES                    | Log output per worker |                      |                        |                  |                       |                      |                        |                  |
|------------------------------|-----------------------|----------------------|------------------------|------------------|-----------------------|----------------------|------------------------|------------------|
|                              | Lower-ranked neighbor |                      | Higher-ranked neighbor |                  | Lower-ranked neighbor |                      | Higher-ranked neighbor |                  |
|                              | OLS                   | IV                   | OLS                    | IV               | OLS                   | IV                   | OLS                    | IV               |
|                              | (1)                   | (2)                  | (3)                    | (4)              | (5)                   | (6)                  | (7)                    | (8)              |
| <i>Log const. expend.</i>    | -0.048***<br>(0.012)  | -0.060***<br>(0.030) | 0.001<br>(0.011)       | -0.004<br>(0.04) | -0.027***<br>(0.008)  | -0.071***<br>(0.030) | 0.001<br>(0.010)       | -0.014<br>(0.05) |
| <i>Neighbor density</i>      | No                    | No                   | No                     | No               | Yes                   | Yes                  | Yes                    | Yes              |
| <i>Angrist-Pischke value</i> |                       | 19.01                |                        | 31.74            |                       | 19.91                |                        | 8.69             |
| Observations                 | 4,088                 | 4,088                | 4,088                  | 4,088            | 4,088                 | 4,088                | 4,088                  | 4,088            |
| R-squared                    | 0.45                  | 0.38                 | 0.45                   | 0.36             | 0.48                  | 0.38                 | 0.46                   | 0.36             |

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Note. Samples include all new entrants, defined as establishments less than 5 years old at the time of the Economic Census, in manufacturing or business services industries responding to Economic Censuses between 1992 and 2007 for which geographic data is available or could be imputed from address records, in tracts neighboring tracts for which there are REITs with positive construction expenditures on new commercial real estate. Neighboring tracts are defined first by ranking all tracts within the CBSA according to the number of establishments per square mile of land area. Neighboring tracts are those with ranks directly above or below tracts with REIT activity, and for which the relative ranks did not switch between any of the four census years. Note that these tracts are not necessarily geographically proximate. All standard errors are clustered at the CBSA level

Columns one of Table 1.8 report the OLS results on less-advantageous locations. Positive shocks to construction shocks have negative effects on the productivity of entrants in the lower-ranked tract, consistent with the sorting hypothesis. Column two uses the same IV as described in Section 4.4.2. Consistent with the findings in the previous subsection, the IV delivers larger negative results. While the density percentiles in Section 4.4.2 likely group large swaths of connected tracts, neighbors in the tract ranking may or may not be physically adjacent. Column one's negative OLS result may be enabled by this divorce from common market conditions like localized demand shocks.

Column three and four replicate the exercise for the next-highest ranked tract. The OLS results are small and positive, consistent with the sorting hypothesis. The IV result in column four results in a noisy, slightly negative coefficient. Although I cannot rule out negative effects, the effects in columns one and three are statistically distinct and their difference is in the direction proposed by the model. Taken together, the results of columns one through four support the sorting hypothesis.

A competing explanation without sorting may be that positive shocks to density in one tract draw some volume of homogenous firms away from other tracts, and the decrease in

density drives the negative relationship. Two pieces of evidence weight against this story. First, because close-ranked tracts may be physically distant, this effect must generate a negative result despite diffusion of the effect of a single construction project on demand across the entire city. Moreover, this effect would operate equally on higher and lower ranked tracts, and would not account for the difference in the coefficients between columns one and three.

Nevertheless, I directly control for changes in establishment density in columns five through eight. Because density is an endogenous variable, these results should be taken lightly. However, the inclusion of observed establishment density directly does not significantly alter the results.

Although the magnitudes of the effects are in line with those found in Section 4.4.2, here I focus on localized effects and a much smaller percent of the overall U.S. geography. I am likely picking up more than 20% of new construction in these roughly one thousand tracts, and the coefficients speak to average effects on a far pool of number of entrants.

Using the model as guidance, the preceding two subsections have isolated positive shocks to density that result in effects on productivity that run counter to the predictions of agglomeration models. Shocks to space provided in the urban core tend to reduce the productivity of entrants relative to incumbents. Shocks to the amount of space provided in a particular location differentially affect other tracts based on their relative advantage. Taken together, these results confirm the suggestive evidence of intra-city firm sorting.

## **1.5 Conclusion**

I have presented a model of heterogeneous firm location decisions where firms trade-off fixed costs of higher rents for increased variable profits, through decreased marginal costs and increased access to markets. Firm sorting and location advantage are co-determined: the sorting pattern interacts with differences in locations' exogenous geographic characteristics to generate endogenous differences in location advantage. Landowners in more advantageous locations are induced to provide more density. I show that in a monopolistic

competition setting with heterogeneous firms where firm location decisions are endogenous, transportation costs to market alone are sufficient to induce the density-productivity relationship. The model both fits the cross-sectional relationships in the data and can be tested against an alternative hypothesis where agglomeration forces alone drive the density-productivity relationship.

The presence of geography often creates tractability issues: it prevents analytic solutions, creates the potential for multiple equilibria, and precludes clear predictions. Faced with this tradeoff, those studying the economic effects of space have largely divorced themselves of the space's theoretical concerns. I introduce a change of variable, indexing the continuous space geography by endogenous location advantage. Location characteristics which are tied directly to geography (in this model, prices and demand) cannot be analytically pinned down, but the remaining location and firm characteristics can. This strategy allows for geographic analysis which neither ignores the role of geography nor succumbs to its complexity. The model is not geographically deterministic but does provide analytic predictions on observables that match the data well. Both these cross-sectional relationships and evidence from establishment moves and firm expansions are consistent with the firm sorting hypothesis.

However, this body of evidence fails to causally distinguish between sorting and agglomeration effects, as does the previous literature on the subject. To that end, the introduction of geography pays dividends. In the model, location advantage is a geographic concept. The composition at firms at each location affect neither the marginal cost of producing at that location, which is defined by the price index, nor the market access of that location, which is defined by a location's geographic proximity to endogenously determined goods markets. In contrast to the literature on inter-city firm sorting, the composition of firms at each location therefore does not affect the sorting behavior of firms. This allows me to test how shocks to certain locations affect firm productivities at other locations in unexpected ways by affecting the pattern of sorting without affecting underlying location fundamentals.

By testing for these composition effects, I find evidence for the existence of intra-city firm

sorting. Examining both the effects of shocks to density in city centers on the productivity of entrants there, as well as the effects of shocks to density at any location on the composition of firms at neighboring locations, I find effects consistent with the sorting hypothesis but inconsistent with models which use agglomeration forces to explain the density-productivity relationship. This constitutes the first reduced-form evidence in favor of the firm sorting hypothesis. To isolate exogenous variation in the supply of space, I develop an instrumental variable based on the inter-city linkages of lessor-developers. Predicting construction expenditures at each location based on the sales and expenditures of single-unit lessor-developers in linked cities, I find predicted expenditures confirm the OLS results and the sorting hypothesis.

While the evidence now weighs in favor of the existence of firm sorting within cities, it does not refute the existence of agglomeration forces such as productivity spillovers. Second, the existent evidence against firm sorting at the inter-city level should be reevaluated within the context of within-city sorting. Combes et al (2012), for example, do not examine whether intra-city heterogeneity can explain the apparent lack of pattern in the productivity cutoffs of firms or the thicker right tail of firm productivity among larger cities. It may well be that sorting within cities replicates the patterns observed in the data.

The flexibility of the location index approach extends beyond its use here. While the model presented in this paper takes a strong stand on the nature of location advantage, the tools I present can be applied more broadly in models where multiple sorting geographic forces are at work. Appendix A5 shows how the strategy can be used to incorporate sorting and agglomeration effects without affecting the static predictions of the model. In future work, the index approach may be used as expressed there in order to quantify the relative strengths of these forces and their interactive effects.

This paper and this new approach bridge two distinct approaches to spatial frictions. Economic geography, which attempts to account for the effects of spatial frictions in a world where locations are networked by their geographic interconnections, and urban economics, where locations are essentially islands that trade freely with each other (or not at all in

certain sectors). Spatial equilibria in urban economics are divorced from actual concerns about proximity, yet equilibrium conditions deliver clear, testable predictions. In economic geography, intractability inhibits models from delivering such clear predictions. This new approach to proximity seeks a middle ground, where geography matters but a broad set of predictions can still be made.

The mechanisms explored in this paper and the empirical results suggest significant new channels through which urban policies may affect urban growth. Restrictions on urban development through zoning regulations may act to change the composition of firm productivities. Zoning regulation, by increasing competition for space, may push out less productive firms.

Two major caveats must be expressed. First, the composition effect operates on the margin. Major changes to supply restrictions affect other general equilibrium channels and can easily overturn the toughened sorting mechanism. Second, this result is dependent on the functional form assumptions of the model. In particular, there are no productivity spillovers in the baseline model. As explored in Appendix A5, the inclusion of productivity spillovers could overturn this result. Nevertheless, the sorting channel should be considered as potentially existent cause of productivity effects of marginal policy changes in zoning restrictions.

In the model, transportation costs are the decisive factor governing market access and therefore the differential density across locations. All roads lead to Rome. However, in this model, decreased transportation costs can also increase market access in the periphery, and push production into the periphery. This double-edged sword, as hypothesized in other geography models (Krugman, 1991), reflects the dichotomy between the opening of the Erie Canal as the impetus for the explosion of density in New York City, and the construction of the suburban highways as a precursor to New York's mid-century urban decline. All roads also lead away from Rome. However, in this model, the flattening of the rent curve is attenuated by the changing composition of firms: as only the most productive firms remain, the cost of the center remains higher than would be predicted in a model of homogenous

firm quality.

Finally, although this paper did not seek to estimate the strength of market access, its predictions reconcile two literatures that have previously done so. Papers in the geography literature such as Ahlfelt *et al.* (2012) have estimated large impacts of changes in market access on the distribution of economic activity. Papers in urban economics that attempt to measure the relative strength of market access such as Ellison and Glaeser (1994) have been more skeptical of the continued importance of location to markets. The results here offer a possible solution: while real differences in market access may be small, they may interact with real estate market elasticities and the sorting behavior of heterogeneous firms to create large differences in the distribution of economic activity.

## Chapter 2

# Unhappy Cities<sup>1</sup>

### 2.1 Introduction

According to the Behavioral Risk Factor Surveillance System (BRFSS), only 35.9 percent of the residents of the Gary, Indiana metropolitan area report themselves as very satisfied with their lives, as opposed to 45.7 percent across the United States as a whole. Self-reported unhappiness is high in other declining cities, and this tendency persists even when we control for income, race and other personal characteristics. Why are the residents of some cities persistently less happy? Given that they are, why do people choose to live in unhappy places?

The presence of significant differences in self-reported well-being across places within the United States poses something of a challenge for the reigning paradigm of urban economics—the concept of a spatial equilibrium. This central idea—proposed by Alonso (1964), Muth (1969), Rosen (1979) and Roback (1982)—assumes that wages and prices adjust so that in equilibrium there are no arbitrage opportunities across space. In equilibrium, individuals cannot improve their overall utility levels by migrating within the U.S.

There are two ways to reconcile differences in self-reported well-being with the notion of a spatial equilibrium. First, subjective well-being (SWB) may not be equivalent to the

---

<sup>1</sup>Co-authored with Edward Glaeser and Joshua Gottlieb

economist's concept of utility. Under this view, agents make decisions in order to jointly maximize expected future happiness and other objectives. Compensating differences in other dimensions offset persistent spatial differences in happiness. Second, the observed differences in subjective well-being may not reflect the permanent life-long well-being for otherwise identical people. The unhappiness might be transitory or explained by unobserved individual heterogeneity, especially if some areas attract people who are disproportionately prone to be more or less happy.

In Section I of this paper, we follow Oswald and Wu (2011) and use BRFSS to measure subjective well-being across the United States. We extend their work by calculating SWB at finer geographic levels, adjusting for observable individual differences, and correcting for sampling error. We find significant, but not huge, differences across metropolitan areas both with and without controlling for state fixed effects. After correcting for sampling noise, we find that the cross-city standard deviation of happiness is about 6 percent of a standard deviation of individual happiness. This is approximately the difference in subjective well-being between the sexes, or between high school graduates and those with some college. This difference is roughly the order of magnitude caused by a one standard deviation decline in neighborhood poverty (Ludwig et al. 2012). We also find that this variation persists when we control for a rich battery of individual controls, including employment status and income.

One primary concern is whether these differences are caused by unobserved heterogeneity, either in human capital or in propensity towards happiness. We address this using the National Survey of Families and Households (NSFH). This is a panel survey, with which we can estimate area-level happiness by looking at individuals who move across metropolitan areas between the survey's first wave (1987-1988) and second wave (1992-1994). Differences in happiness persist, even when we control for individual fixed effects. The correlation between area level estimates with and without individual fixed effects is 0.69. This leads us to believe that much of the difference in happiness across space reflects more than the selection of unhappy people into unhappy places.



We next document that area-level happiness is essentially uncorrelated with many area attributes. For example, metropolitan area population and housing values are orthogonal to subjective well-being in the BRFSS. Like Florida et al. (2013), we find that area-level education is positively associated with subjective well-being, but we find that this effect vanishes when we control for individual-level education. If more educated individuals only became educated because of the education level of the area, then it can be fairly said that these places have made them happier. But if they would have been educated regardless of place, then the happiness of more educated areas should be interpreted as differential selection.

In Section II, we document the one robust fact that emerges clearly from multiple data sets: places with lower levels of population and income growth are less happy (Glaeser and Redlick, 2009). Lucas (2013) also finds higher rates of migration to counties with higher subjective well-being in the BRFSS, arguing that the migration patterns are consistent with a spatial equilibrium with happiness as a measure of utility. We find the relationship persists for quite long periods (from 1950 to 2000). Moreover, we find the strongest effect at the left tail of SWB. It is not that high-growth places are particularly happy, but rather that very low-growth areas are particularly unhappy. It is possible that people flee areas that produce unhappiness, but the long time periods involved make it hard to believe that these differences are transitory.

We show that the connection between low well-being and decline persists when we control for a bevy of individual controls, including education and income, and even when we control for state fixed effects. This fact appears in three independent surveys. In the NSFH, the effect does not persist in the general individual fixed effects estimation, but it re-emerges when we limit our sample to cities with more than 250 respondents across both waves. None of these results speak to whether unhappiness is causing decline or whether decline is causing unhappiness.

Section II also notes three other facts about urban decline and unhappiness. First, while Oswald and Wu (2010) document the relationship between state-level happiness and

amenities, we find that the connection between unhappiness and decline in the BRFSS does not reflect the role of urban disamenities associated with decline, such as crime, coldness and inequality. Second, we find that the connection between urban decline and low SWB is just as strong among recent migrants as among longer-term residents. This latter fact leans against the interpretation that happiness was *ex ante* identical across areas, but that some areas experienced negative shocks, people were stuck in those areas and their happiness fell accordingly.

Third, we ask whether the unhappiness of declining cities is a new phenomenon, perhaps caused by decline, or represents a more historic tendency. The General Social Survey (GSS) enables us to look back as far as the early 1970s, and these data suggests that the connection between decline and unhappiness was stronger in the past than it is today. These facts lead us to suspect that the connection with unhappiness and urban decline more likely reflects long-standing attributes of these cities rather than a causal effect of the decline itself.

In Section III, we propose a framework that incorporates spatial differences in SWB into the spatial equilibrium framework. Following writers as diverse as Epictetus, de Mandeville, Irving Fisher and Gary Becker, we assume that happiness is desirable—but not equivalent to utility. We have life objectives other than being satisfied, and may knowingly make choices that reduce happiness, such as exposing ourselves to a more competitive environment, if those choices further other aims (Luttmer 2005; Benjamin et al. 2011). According to the spatial equilibrium logic, a city's unhappiness must be offset by some other amenity, such as higher real income.

In our model, happiness is generated through experiences, which can be improved by spending money, and happiness is but one ingredient in the utility function. Individuals have other objectives, which we refer to as achievements, such as raising a family. These are also produced with a combination of money and time. The model suggests that the connection between money and happiness may significantly understate the connection between money and utility, because a higher opportunity cost of time causes individuals to engage in less happiness-generating leisure. In a spatial equilibrium, higher wages are

compensated shifts, typically offset by higher real estate prices, so higher area wages could easily be associated with lower happiness levels even if utility levels are equalized across space.

In Section IV, we examine whether individuals in declining or otherwise unhappy places are being compensated for their unhappiness. In the 1940 Census, residents of declining cities were receiving significantly higher incomes. A one standard deviation drop in population growth post 1950 was associated with 222 more in income (3,655 in current dollars), which is more than ten percent of average income. Presumably, high labor costs were one reason why businesses left these areas. One interpretation of these results is that the industrial cities were less happy in 1940, but their residents were being compensated with earnings that could achieve other ends, such as nurturing a family.

The data also shows that housing prices in 1940 were higher in areas that subsequently declined, yet there are essentially no housing quality controls in that early data. As such, while it is possible that some of the high earnings in declining cities were eaten away by higher housing rents, it is also possible that these rents were actually compensation for better housing quality.

When we turn to 2000 Census data, we find that the unhappy, declining cities are no longer receiving higher wages. Wages are essentially uncorrelated with our growth variable in the more modern data. But decline is correlated with house prices and rents. In 1940, the residents of unhappy, declining places seem to have been compensated with higher incomes. In 2000, the residents of those same cities seem to have been compensated with lower housing costs.

We have also examined the direct correlation between our area-level happiness measure and area-level rents and incomes, as in Oswald and Wu (2011). We do find some evidence that residents of happy cities pay higher rents, suggesting some form of offset for the added level of happiness. The results are certainly compatible with the view that individuals trade other objectives against happiness when they are choosing where to live. Section VI concludes.

## 2.2 Unhappiness Across Cities

To begin, we briefly document five stylized facts about urban happiness, primarily in the U.S., but also abroad. We discuss the connection between unhappiness and decline in Section III.

Throughout this paper, we follow the literature in measuring happiness using self-reported survey data on subjective well-being (SWB). Our primary data source is a large national survey, the Behavioral Risk Factor Surveillance System (BRFSS) conducted by the Centers for Disease Control and Prevention (CDC), which asks individuals to report on their own life satisfaction using a discrete response scale.

Since 2005, CDC has asked all respondents “In general, how satisfied are you with your life?” Respondents were given four possible categories: very satisfied, satisfied, dissatisfied, and very dissatisfied. In each year between 2005 and 2010, around 300,000 subjects answer this question, along with all of the demographic variables listed below.<sup>2</sup> This question has been the focus of much of the previous literature on the economics of happiness, and we show the distribution of answers in Appendix Table 1. We recognize that satisfaction may strictly differ from happiness, but we will use the terms interchangeably.

In all of the work that follows, we recode these answers so that 4 indicates “very satisfied” and 1 indicates “very dissatisfied.” We then rescale the answers linearly so that they have a mean of 0 and standard deviation of 1. Because BRFSS reports the county in which the respondent lives, we are able to link respondents to metropolitan areas.

This measure has several problems, even before considering whether it corresponds to the economic concept of “utility.” First, respondents may have different interpretations of the response scale, or equivalently different reference points for life satisfaction. A situation that one person may consider very satisfactory, may be merely satisfactory to another. If this leads to systematic differences across individuals, it could confound the variable’s

---

<sup>2</sup>We discuss this survey in more detail in the Data Appendix. We also explore the issues that arise from the discrete nature of the answers and explain why we do not think they are a problem, as well as other details of our estimation.

interpretation.

To address this, we estimate area  $j$  happiness as the MSA  $j$  fixed effect in the following model:

$$y_{ijt} = \alpha + X_{ijt}\beta + \gamma_t + u_j + \epsilon_{ij} \quad (2.1)$$

We estimate equation (1) at the individual level, so  $i$  indexes individual respondents,  $j$  indexes areas, and  $t$  indexes the survey wave. In this regression,  $y_{ijt}$  represents individual subjective well-being (or SWB),  $X_{ijt}$  is a matrix of individual controls,  $u_j$  is a metropolitan area fixed effect,  $\gamma_t$  is a year fixed effect, and  $\epsilon_{ij}$  is an uncorrelated error term. The individual controls include survey month, sex, a polynomial in age, eight race dummies, six marital status dummies, four educational attainment dummies, and variables representing various information about the children in the household. See the Data Appendix for more details on these controls.

Second, respondents undoubtedly have a large degree of variability in their happiness at the moment they answer the survey. Because we only have responses from a small fraction of residents in each area (around 0.1%), this variability is likely to cause noisy estimates of area-level SWB. To account for this, we next measure area-level happiness using random effects instead of fixed effects. We estimate the following model, in which coefficients in bold type are considered to be fixed, while the others are random effects:<sup>3</sup>

$$y_{ijt} = \alpha + X_{ijt}\beta + \gamma_t + u_j + \epsilon_{ij} \quad (2.2)$$

We consider the demographic characteristics to have a fixed relationship with individual happiness, and allow for random metropolitan area effects as well as an individual error term.

---

<sup>3</sup>So one might prefer to call this a “mixed effects” model as opposed to a pure “random effects” model.

This model enables us to compute a number of useful quantities. It allows us to calculate an estimate of the underlying variance of metropolitan area effects ( $\sigma_u^2$ ). For each area, we can also determine the best estimate  $\hat{u}_j$  of that area's  $u_j$ . We refer to these estimates as the metropolitan area's adjusted life satisfaction. We use them extensively in subsequent analysis as our estimate of the area's contribution to individual happiness.<sup>4</sup>

Finally, since BRFSS has only asked about life satisfaction since 2005, we have a limited ability to address time-series variation in happiness. We will thus augment it with other data sources introduced below. We first turn to five sets of facts about life satisfaction across space.

### 2.2.1 Are There Significant Differences in Life Satisfaction across Space?

We first address whether there is a meaningful difference in happiness levels across geographic areas, both before controlling for individual demographic characteristics and after including these controls. We answer this question in multiple ways. First, we run the fixed effects regression (1) and perform an F-test of the joint significance of the metropolitan area fixed effects. Second, we determine whether the estimated variance of metropolitan area random effects in regression (2),  $\sigma_u^2$ , is significantly different from zero. Third, we perform a likelihood ratio test of the fixed effects model (2) against a constrained model in which the random effects are removed (we force  $u_j = 0$  for all  $j$ ).

We run each of these tests on a model with no demographic controls, and with the full set of demographic controls shown in Appendix Table 2. In both cases, all three tests strongly reject the null hypothesis that metropolitan area effects are irrelevant, and all with  $p < 0.0001$ .

Our next task is to quantify the differences across regions. We do so using two different measures from the random effects estimates in (2). First,  $\sigma_u^2$  provides an estimate of the

---

<sup>4</sup>Our calculation of these adjusted life satisfaction measures  $\hat{u}_j$  recognizes the problem of potentially large sampling variation when measuring SWB in a survey. We therefore calculate the best linear unbiased predictor (BLUP) based on our MSA-level random effects from (2), following the method of Bates and Pinheiro (1998) as implemented in Stata 11.1.

variance across the full population of metropolitan and non-metropolitan areas. Second, the empirical variance of the adjusted life satisfaction values,  $\text{Var}(\hat{u}_j)$ , quantifies the dispersion of estimates in the sample of areas where we are able to compute happiness.

In the unadjusted random effects model (where  $X_{ijt}$  is empty so we have no demographic estimates  $\beta$ ), we find  $\sigma_u = 0.063 \pm 0.004$  and  $\text{sd}(\hat{u}_j) = 0.058$ . Since all of our analyses use measures of SWB rescaled to have zero mean and unit variance across individuals, the variation across geographic regions is around 6% of the individual-level variation in happiness. These numbers shrink by about one-quarter, to  $\sigma_u = 0.047 \pm 0.003$  and  $\text{sd}(\hat{u}_j) = 0.042$ , when we include the demographic controls in model (2). Appendix Figure 1 shows the distribution of these adjusted life satisfaction estimates.

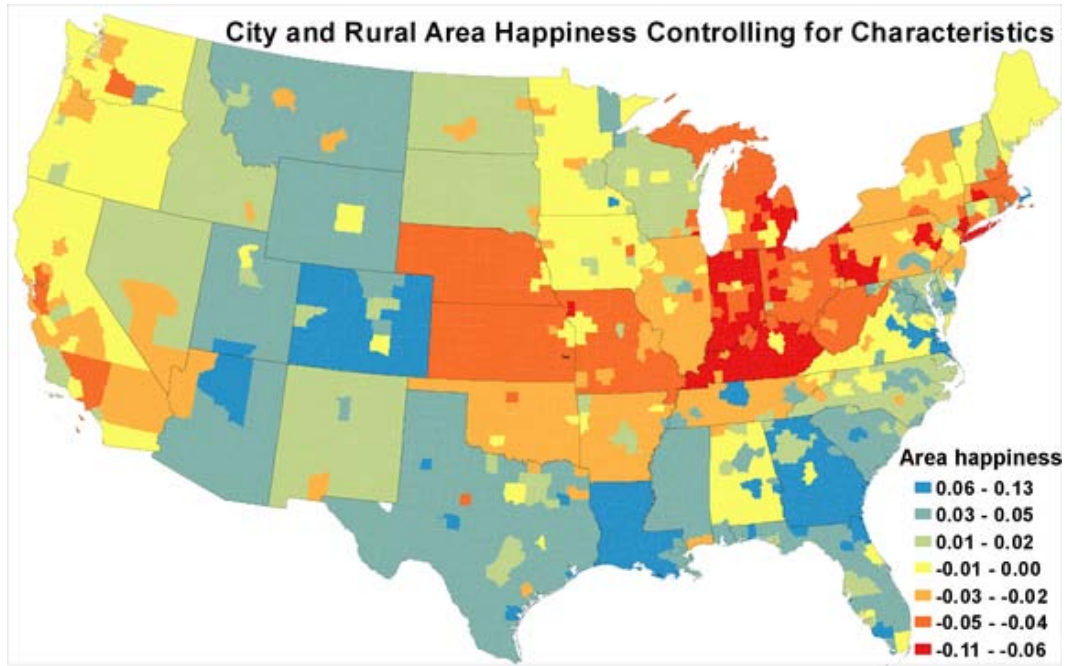
To get a better sense of what this means quantitatively, we can compare it to the estimates of the impact of other characteristics on individual SWB. For example, moving across one standard deviation in geographic areas has an impact one-third as large as the difference between being a high school graduate or not graduating, or 1.8 times the estimated male-female gap.

The values of our local happiness estimates themselves are shown visually in Figure 2.1. This map shows adjusted life satisfaction estimated at the MSA and rural area level after controlling for individual demographics.<sup>5</sup> The map shows a band of less happy areas in parts of the Midwest and the Appalachian states, stretching from Missouri in the west and Alabama in the south well into Pennsylvania and even New Jersey in the east. New York City, Detroit, and much of California also have lower SWB, while the happiest areas are concentrated in the West, Upper Midwest, and rural South. Appendix Table 3 shows specific values for a handful of metropolitan and non-metropolitan regions, including the highest and lowest values that we estimate.

A third potential problem with these results is that they may reflect differences in the ways in which states implement the BRFSS. Unlike many surveys, the BRFSS is not centrally

---

<sup>5</sup>Because income depends on numerous individual choices, including where to live and possibly including happiness, we do not include it among our demographic controls. For the interested reader, our working paper version (Glaeser, Gottlieb and Ziv 2014) shows a version of this map after adjusting for individual income.



**Figure 2.1:** *Estimated Metropolitan and Rural Area Adjusted Happiness*

administered. Instead, individual state agencies perform the surveys. We cannot be sure of what biases may be created through this decentralized implementation, but it is at least possible that state-level implementation has caused some of the variance that we see in the data.

To address this possibility we re-estimate model (2) controlling for state fixed effects. Since there are a relatively few number of metropolitan areas in many states, we will not use these state-corrected area fixed effects in general. Still, it is important to note the reduction in variance that occurs when we look only at the within-state variance. The standard deviation of  $\hat{u}_j$  falls from 0.043 to 0.017 when we control for state fixed effects as well as demographic controls. The variance is significantly reduced, but these effects remain statistically distinct from zero. As such, we conclude that metropolitan differences would persist, even if all the state-level variation reflected only state-level differences in implementing the BRFSS.

This evidence does not rule out the possibility that these differences reflect unobserved individual characteristics. One approach to unobserved heterogeneity is to estimate



metropolitan area fixed effects controlling for individual fixed effects. This requires us to use a panel, rather than a repeated cross-section, which forces us to move from the very large BRFSS to the much smaller National Survey of Families and Households (NSFH). The NSFH is a longitudinal study, from which we use the first two waves (completed in 1987-1988 and 1992-1994) (Sweet and Bumpass 1996; Sweet, Bumpass, and Call 1988; Trull and Famularo 1996).

In both waves, the data contains information on family and personal characteristics of individuals and on individual subjective well-being. In particular, the NSFH asks: “First taking things all together, how would you say things are these days?” Appendix Table 1 shows the responses.

We will later use this measure to examine whether the link between area attributes and well-being is stronger for recent migrants or long-term residents. Here, we restrict our attention to the heterogeneity in subjective well-being across space. We first estimate adjusted life satisfaction for the merged sample of NSFH waves 1 and 2. The variance of these estimates is 0.0007, roughly in line with the estimates from the BRFSS. The raw variation of metropolitan area fixed effects is larger in the NSFH, but the variance correction is also much larger because the sample size is so much smaller.

We then estimate a PMSA fixed effect variable using the two waves including individual level fixed effects. The correlation between these estimates and the estimates without the individual fixed effects is 0.69. The variance of the PMSA fixed effect with individual fixed effects is 0.64. We conclude from these results that there appears to be significant variation in subjective well-being across space, even when we control for unobservable individual-level heterogeneity by using individual fixed effect estimates.

### **2.2.2 Do Metropolitan Area Differences in Subjective Well-Being Persist?**

Having established the existence of spatial differences in happiness and estimated their magnitude, we now want to see how they evolve over time. Hypotheses about the temporal pattern of spatial SWB could range from a completely permanent local characteristic (for

instance, Honolulu has gorgeous weather and is on the beach, which always makes its residents happy) to a long-term shock common to area-level residents (e.g., the economy in Detroit was poor and declining during our sample period, making its residents unsatisfied), to an extremely transitory common shock caused by the weather or local sports outcomes.

We first test the stability of area effects in two ways. First, we run versions of regression (2) separately for each year, so without year fixed effects:

$$y_{ij} = \alpha^{(t)} + X_{ij}\beta^{(t)} + u_j^{(t)} + \epsilon_{ij} \quad (2.3)$$

We then compare the adjusted life satisfaction estimates across different years ( $\hat{u}_j^{(t)}$ ) versus  $\hat{u}_j^{(t')}$  for  $t' \neq t$ ). Our second method is to augment regression (2) by adding an area-year random effect,  $v_{jt}$  to the random effects regression:

$$y_{ijt} = \alpha + X_{ijt}\beta + \gamma_t + u_j + v_{jt} + \epsilon_{ij} \quad (2.4)$$

This model estimates the time-invariant area effect,  $u_j$  and the time-varying area effect  $v_{jt}$  simultaneously. We can test the statistical impact of each of these effects separately, and quantify the importance of permanent and transitory area effects. For this analysis, we use the sample of respondents in the 177 MSAs with at least 200 respondents in all years of our sample.

These tests reveal very clearly that the permanent effects are far more important than the transitory components. When estimating equation (4) without demographic controls, we find  $\sigma_u = 0.064 \pm 0.004$  while  $\sigma_v = 0.018 \pm 0.002$ . Thus there is a statistically significant transitory component, but it varies by 70% less than the permanent area component, and its standard deviation is around 2% of the individual-level standard deviation.

Another way to see this variation is to relate adjusted life satisfaction from regression (3) in one year to that in another year. Using the measures adjusted for demographic controls from 2005 ( $\hat{u}_j^{(2005)}$ ) and from 2009 ( $\hat{u}_j^{(2009)}$ ), we find an extremely strong positive

relationship, with a correlation of 0.48. Thus one quarter of the variation in adjusted life satisfaction is driven by permanent metropolitan area level shocks, and the rest by transitory shocks and estimation error. Although we have adjusted for the effect of sampling error in computing adjusted life satisfaction, we should expect to see a correlation less than one if our correction is imperfect. Hence the random effects results discussed in the previous paragraph give the more accurate assessment of the relative importance of permanent and transitory components to well-being.

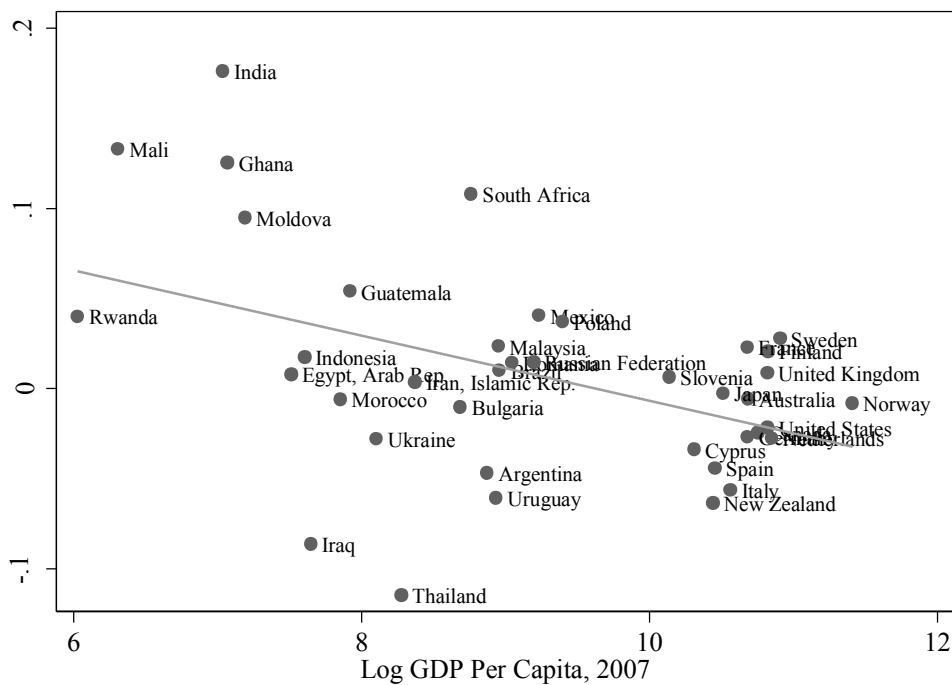
### **2.2.3 Is Urbanization Associated with Happiness or Unhappiness?**

One natural question is whether happiness increases or diminishes in large cities. Cities have often been seen as entities that create financial wealth, but diminish other types of well-being. We first test this hypothesis by examining the correlation between adjusted life satisfaction and the logarithm of metropolitan area population. If we use the 2010 population, we find a weak positive correlation of 0.07. As metropolitan area population increases by one log point, SWB increases by 0.003 standard deviations and the effect is quite imprecise. We will later show that in an individual level regression, there is also no significant relationship between area level population and self-reported well-being, holding individual level characteristics constant.

Using past population levels, instead of current levels, we find a positive correlation with population in 2000 and 1990 and a negative correlation with population levels before that point. The relationship between recent levels of SWB and metropolitan area population before 1960 becomes significantly negative. While larger cities today do not evince significantly lower levels of unhappiness, residents of cities that were large in the past do seem to be less happy. We return to this topic later, when we discuss the connection between SWB and population growth.

Following Stevenson and Wolfers (2008), we now briefly turn to worldwide data. Using the World Values Survey, we estimate subjected well-being in rural and urban areas in 39 countries throughout the world. Across the entire sample, we find that the urban happiness

is on average higher than rural happiness. This effect is, however, driven primarily by poorer countries.



**Figure 2.2:** *Urban-rural happiness gradient vs GDP across countries*

Figure 2.2 shows the correlation between the logarithm of per capita GDP in the country in 2007 and the rural-urban gap in subjective well-being. The coefficient is significantly negative and the  $R^2$  is 0.2. In poorer countries, which often have cities that seem particularly hellish, the residents of cities say that they are significantly happier than the residents of rural areas. It is perhaps unsurprising that the developing world is urbanizing so rapidly, as urban residents appear to be both far better paid and happier.

## 2.2.4 Unhappiness and Urban Characteristics

We now turn to area-level correlates of self-reported well-being. Most area level attributes are relatively uncorrelated with subjective well-being, at least conditional on individual education.

Table 2.1 presents these facts using area characteristics as of the year 2000. We use the

year 2000 both because it predates our well-being data and because it is the last year with a comprehensive census. Our core specification includes a bevy of individual attributes that have been found to correlate with happiness, including education, age, race and family status. We do not include income or employment controls as these represent outcomes that may be caused by an area's economic success. Education and marital status may themselves be determined by the urban environment, and we include regressions both with and without those controls. All regressions cluster the standard errors at the area-year level and include year and month fixed effects.

**Table 2.1: Happiness levels across space, BRFSS**

| Dependent variable:       | (1)                      | (2)                 | (3)                 | (4)              | (5)                 | (6)              | (7)                  | (8)                  | (9)                  | (10)                 |
|---------------------------|--------------------------|---------------------|---------------------|------------------|---------------------|------------------|----------------------|----------------------|----------------------|----------------------|
|                           | Self-Reported Well-Being |                     |                     |                  |                     |                  |                      |                      |                      |                      |
| Log Population, 2000      | -0.007<br>(0.006)        | -0.009**<br>(0.004) |                     |                  |                     |                  |                      |                      | -0.003<br>(0.004)    | 0.001<br>(0.003)     |
| % BA Grad, 2000           |                          |                     | 0.330***<br>(0.108) | 0.098<br>(0.077) |                     |                  |                      |                      | 0.394***<br>(0.114)  | 0.185***<br>(0.062)  |
| % HS Grad, 2000           |                          |                     |                     |                  | 0.421***<br>(0.151) | 0.072<br>(0.116) |                      |                      | -0.276**<br>(0.125)  | 0.089<br>(0.083)     |
| Segregation Index, 2000   |                          |                     |                     |                  |                     |                  | -0.160***<br>(0.038) | -0.130***<br>(0.03)  | -0.130***<br>(0.034) | -0.032<br>(0.033)    |
| Segregation x Black       |                          |                     |                     |                  |                     |                  | -0.263***<br>(0.051) | -0.144***<br>(0.048) | -0.113**<br>(0.044)  | -0.057<br>(0.039)    |
| Segregation x Asian       |                          |                     |                     |                  |                     |                  |                      |                      | 0.051<br>(0.088)     | 0.0872<br>(0.058)    |
| Segregation x HPI         |                          |                     |                     |                  |                     |                  |                      |                      | -0.037<br>(0.151)    | -0.049<br>(0.154)    |
| Segregation x Other       |                          |                     |                     |                  |                     |                  |                      |                      | 0.217**<br>(0.101)   | 0.230**<br>(0.101)   |
| Segregation x AIAN        |                          |                     |                     |                  |                     |                  |                      |                      | -0.070<br>(0.083)    | -0.058<br>(0.082)    |
| Segregation x Multiracial |                          |                     |                     |                  |                     |                  |                      |                      | -0.160***<br>(0.062) | -0.140***<br>(0.053) |
| Segregation x Hispanic    |                          |                     |                     |                  |                     |                  |                      |                      | -0.055<br>(0.247)    | -0.252<br>(0.205)    |
| Log Median House Value    |                          |                     |                     |                  |                     |                  |                      |                      | -0.040<br>(0.026)    | -0.045***<br>(0.011) |
| Additional Controls       | No                       | Yes                 | No                  | Yes              | No                  | Yes              | No                   | Yes                  | Yes                  | Yes                  |
| State Fixed Effects       | No                       | No                  | No                  | No               | No                  | No               | No                   | No                   | No                   | Yes                  |
| Observations (Thousands)  | 1,185                    | 1,185               | 1,185               | 1,185            | 1,185               | 1,185            | 1,134                | 1,134                | 1,134                | 1,134                |
| R <sup>2</sup>            | 0.007                    | 0.076               | 0.008               | 0.076            | 0.008               | 0.076            | 0.008                | 0.076                | 0.076                | 0.078                |

Sources: Authors' regressions on microdata from the Behavioral Risk Factor Surveillance System Survey (CDC), U.S. Census (Ruggles et al. 2010), and Glaeser and Vigdor (2001).

Notes: All regressions control for year fixed effects, month fixed effects, age, race, and sex. "Additional controls" include education, marital status, and family size. Standard errors in parentheses are clustered at the MSA level (\*\*\*)  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Our first regression shows the relationship between the population size of the metropolitan area and self-reported well-being. When we do not control for education and marital status, the statistical relationship is small and statistically indistinct from zero. When we include these more endogenous controls, the relationship becomes more negative and statistically significant.

In the third regression, we control for the share of the adult population in the area with

a college degree. Using the fixed effects estimated without controlling for individual level education, this variable is strongly positive. Using the fixed effects estimated conditional on these controls, the variable's estimated effect drops by two-thirds and it becomes statistically indistinct from zero. Regressions five and six examine the share of the adult population with a college degree. The picture is much the same as with the other education variable. The coefficient is large and statistically significant when we control only for area level attributes but not when we control for area-level education. These regressions can be interpreted as suggesting that area level education boosts self-reported well-being by increasing individual educational attainment, or that area level education has no independent effect.

Regressions seven and eight examine racial segregation, as measured by a standard dissimilarity index. In this case, we also interact segregation with a dummy variable that takes on a value of one if the individual is black. Both with and without individual controls, segregation is negatively associated with well-being and this effect is approximately twice as large for African-Americans as for whites.

Regression nine controls for all of the metropolitan area variables, and the full set of individual level controls. We also interact segregation with all of the race categories. The results here remain similar to our previous specifications. The population share with a college degree has a positive effect on self-reported happiness, although the share with a high school degree has a negative relationship. Segregation continues to have a negative connection to self-reported well-being, and this effect is much stronger for African-Americans. In this specification, housing value has a somewhat surprisingly negative but insignificant effect on subjective well-being.

Column ten adds state fixed effects and is our most complete specification. As many states have only one metropolitan area, this reduces our effective sample and eliminates any variation that represents larger regional trends. In this specification, the positive effect of college education remains, and the effect of high school graduates becomes positive. Housing values now have a negative, significant, relationship with happiness, while segregation is insignificant.

Putting together these results, we draw two tentative conclusions. There is some possibility that individuals report higher levels of well-being in more educated areas, although this is true only when we include a full range of area controls, or when we fail to control for individual level education. Segregation is associated with lower levels of subjective well-being, but only when we don't control for state fixed effects. Overall, these results do not suggest a robust series of correlations between urban attributes and SWB.

## **2.3 Unhappiness and Urban Decline**

We now turn to a particularly striking correlation between urban unhappiness and decline (Glaeser and Redlick 2009, Lucas 2013). We first examine linear specifications and then allow the impact of population growth on subjective well-being to have a piecewise linear shape. We will focus on changes in the logarithms of population and median household income between 1950 and 2000. We first focus on the BRFSS and then turn to the NSFH and the GSS, which enable us to look at movers and estimate equations with individual fixed effects.

### **2.3.1 Linear Effects of Population Growth and Income**

Table 2.2 presents our first set of results on the correlation between SWB and urban change. The first three regressions show results for population change. The next regression shows results for income change. The final two regressions show results for both variables together and include other area-level controls.

The first regression shows the relationship between population change and self-reported well-being controlling for individual attributes. The coefficient of 0.0635 implies that a 1 log point increase in population growth is associated with about a one-sixteenth of a standard deviation increase in self-reported well-being. The second regression controls for the more endogenous individual characteristics. The coefficient on population change remains statistically significant, but it falls in magnitude by about one-third. The third regression controls for state fixed effects. In this case, the coefficient falls to about one-fourth of its value in the first regression, although it retains statistical significance.



**Table 2.2: Happiness and urban characteristics, BRFSS**

| VARIABLES                           | (1)                   | (2)                    | (3)                      | (4)                   | (5)                    | (6)                    |
|-------------------------------------|-----------------------|------------------------|--------------------------|-----------------------|------------------------|------------------------|
|                                     |                       |                        | Self-Reported Well-Being |                       |                        |                        |
| Change in Log Population, 1950-2000 | 0.0635***<br>(0.0146) | 0.0412***<br>(0.00925) | 0.0174***<br>(0.00649)   |                       | 0.0270***<br>(0.0104)  | 0.00312<br>(0.00710)   |
| Change in Log Income, 1950-2000     |                       |                        |                          | 0.0586***<br>(0.0156) | 0.0597**<br>(0.0298)   | 0.0301<br>(0.0216)     |
| % BA Grad, 2000                     |                       |                        |                          |                       | 0.124<br>(0.113)       | -0.0260<br>(0.0406)    |
| % HS Grad, 2000                     |                       |                        |                          |                       | -0.170<br>(0.140)      | 0.185***<br>(0.0706)   |
| Segregation Index, 2000             |                       |                        |                          |                       | -0.0930***<br>(0.0230) | -0.0484***<br>(0.0180) |
| Additional Controls                 | No                    | Yes                    | Yes                      | Yes                   | Yes                    | Yes                    |
| State Fixed Effects                 | No                    | No                     | Yes                      | Yes                   | No                     | Yes                    |
| Observations                        | 1,182,563             | 1,182,563              | 1,182,563                | 1,166,056             | 1,114,898              | 1,114,898              |
| R-squared                           | 0.008                 | 0.076                  | 0.078                    | 0.078                 | 0.077                  | 0.078                  |

Sources: Authors' regressions on microdata from the Behavioral Risk Factor Surveillance System Survey (CDC), U.S. Census (Ruggles et al., 2010), and GI

Notes: All regressions control for year fixed effects, month fixed effects, age, race, and sex. "Additional controls" include education, marital status, and family size. Standard errors in parentheses are clustered at the MSA level (\*\*\* p<0.01, \*\* p<0.05, \* p<0.1).

Regression four looks at income change instead of population change, including all of the same controls as column three. In a sense, this is the local version of the classic Easterlin (1974) work on income change and happiness. It shows a strong positive relationship between income growth and self-reported well-being. The coefficient is somewhat larger than that on population growth, but since the variation of income growth is smaller, the impact of a one-standard deviation change in income growth is actually smaller than the impact of a one-standard deviation change in population growth.

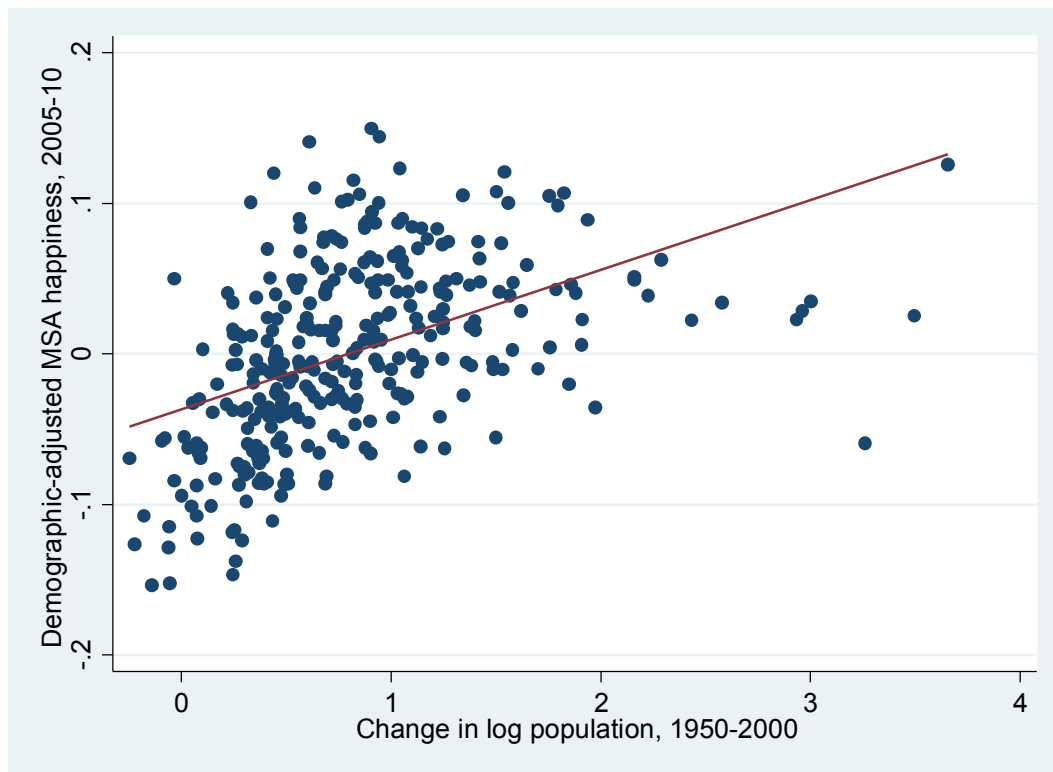
Regression five includes both change variables and other area level controls, and the change variables both remain statistically significant. The coefficients are modest but continue to suggest that growth is associated with positive levels of well-being. The only other control that is statistically distinct from zero is segregation, which remains negative.

As BRFSS is administered at the state level, there could be cross-state differences in the survey's implementation. To adjust for any such differences, we add state fixed effects in regression six. Note that the fixed effects may be over-controlling in important ways, as they eliminate all regional variation from our estimates. Our map of adjusted happiness

(Figure 2.1) shows clear regional patterns, and we now eliminate that variation and more. With this caveat, column 6 shows that these fixed effects eliminate the otherwise robust relationship between urban growth and subjective well-being. But unfortunately they do not tell us whether this reflects variability in survey implementation or is simply because the bulk of geographic differences in subjective well-being are regional in nature.

As we will see in the next sub-section, the growth-happiness relationship is driven by the lower end of the city growth distribution. This part of the relationship remains robust to state fixed effects, reducing the importance of distinguishing between columns 5 and 6 of Table 2.2.

### 2.3.2 The Non-Linear Relationship between Happiness and Population Growth



**Figure 2.3:** *Estimated Metropolitan and Rural Area Adjusted Happiness*

Figure 2.3 shows the correlation between population growth and adjusted life satisfaction.

As the figure makes clear, the effect is much stronger at the lower end of the population change distribution. Low levels of happiness are particularly common in areas with declining population, but higher levels of happiness are not especially prevalent in areas where population is growing rapidly.

There are several hypotheses that could explain this non-linearity. For example, if decline is actually causing unhappiness—rather than merely being correlated with it—it might be that decline itself creates urban stresses, relative to stasis, but that urban growth doesn't particularly alleviate those stresses. Declining cities, such as Detroit, often find it difficult to cover the costs of their historic footprint and infrastructure. Decline may be associated with crumbling social or physical infrastructure. It could also be that happiness is caused by other attributes that cause decline, but that among growing cities, the differences come mainly from differences in housing supply and economic productivity, which perhaps have little impact on happiness.

Whatever the cause, the non-linear relationship is obvious in the data. Table 2.3 shows the connection between SWB and urban growth in the BRFSS, where we have allowed the break in the slope to occur at a value of 0.75, the median for our sample of metropolitan areas. Column 1 shows that, controlling for exogenous demographic controls, the coefficient on growth when growth is below the median is 0.214, meaning that a 0.5 change in log population growth is associated with a 0.1 standard deviation increase in SWB. The result is extremely significant and remains so in the second regression, where we include the endogenous demographic controls. The coefficient here drops to 0.134, meaning that a 0.5 change in log population growth is associated with a 0.065 standard deviation increase in SWB. This is roughly equivalent to one standard deviation of the metropolitan area fixed effects, and roughly equivalent to the difference in SWB between high school graduates and individuals who have some college education.

In regression 3, we control for income and employment status. While we recognize endogeneity of these outcomes with respect to local labor market conditions, we still think it is worthwhile knowing whether the connection between urban decline and unhappiness

**Table 2.3: Happiness and urban population growth differences, BRFSS**

| Dependent variable:                               | (1)                  | (2)                  | (3)                  | (4)                      | (5)                   | (6)                     | (7)                      | (8)                     |
|---|----------------------|----------------------|----------------------|--------------------------|-----------------------|-------------------------|--------------------------|-------------------------|
|   |                      |                      |                      | Self-Reported Well-Being |                       |                         |                          |                         |
| Change in log population (below median) 1950-2000 | 0.214***<br>(0.0186) | 0.134***<br>(0.0146) | 0.101***<br>(0.0174) | 0.0972***<br>(0.0180)    | 0.0781***<br>(0.0127) | 0.156***<br>(0.0344)    | 0.0863***<br>(0.0251)    | 0.0830*<br>(0.0446)     |
| Change in log population (above median) 1950-2000 | 0.00409<br>(0.0127)  | 0.00503<br>(0.00795) | 0.00929<br>(0.00711) | 0.00771<br>(0.00642)     | -0.00443<br>(0.00564) | 0.0231<br>(0.0331)      | 0.0134<br>(0.0224)       | 0.0164<br>(0.0240)      |
| Average January temperature                       |                      |                      |                      |                          |                       | 0.00153**<br>(0.000780) | 0.00177***<br>(0.000442) | -0.00281**<br>(0.00134) |
| Precipitation                                     |                      |                      |                      |                          |                       | 0.000117<br>(0.000682)  | -0.000146<br>(0.000480)  | 0.00190**<br>(0.000840) |
| Log of Crime                                      |                      |                      |                      |                          |                       | 0.00457<br>(0.0115)     | 0.00258<br>(0.00703)     | 0.0115<br>(0.00707)     |
| Pollution   |                      |                      |                      |                          |                       | 0.000171<br>(0.00140)   | 0.000487<br>(0.000955)   | -3.81e-05<br>(0.00107)  |
| Gini coefficient, 2000                            |                      |                      |                      |                          |                       | -0.0538<br>(0.643)      | 0.325<br>(0.420)         | 0.863***<br>(0.237)     |
| Additional Controls                               | No                   | Yes                  | Yes                  | Yes                      | Yes                   | No                      | Yes                      | Yes                     |
| Employment and Income Controls                    | No                   | No                   | Yes                  | Yes                      | No                    | No                      | No                       | No                      |
| Health Controls                                   | No                   | No                   | No                   | Yes                      | No                    | No                      | No                       | No                      |
| State Fixed Effects                               | No                   | No                   | No                   | No                       | Yes                   | No                      | No                       | Yes                     |
| Observations                                      | 1,182,563            | 1,182,563            | 1,182,563            | 1,164,203                | 1,182,563             | 261,987                 | 261,987                  | 261,987                 |
| R <sup>2</sup>                                    | 0.009                | 0.077                | 0.125                | 0.185                    | 0.078                 | 0.010                   | 0.078                    | 0.079                   |

Source: Authors' regressions on microdata from the Behavioral Risk Factor Surveillance System Survey (CDC) and U.S. Census (Ruggles et al., 2010).

Notes: All regressions control for year fixed effects, month fixed effects, age, race, and sex. "Additional controls" include education, marital status, and family size.

Standard errors in parentheses are two-way clustered (Cameron, Gelbach and Miller 2011) at both the MSA and year levels (\*\*\*) p<0.01, \*\* p<0.05, \* p<0.1).

disappears when we control for them. In this case, the coefficient falls to 0.101. In regression four, we control for health status (including both a general question about overall health status and a question about days spent ill over the past year). The coefficient falls to 0.097 and remains quite statistically significant. Column 5 adds state fixed effects, but excludes the health and income questions. Here we estimate a coefficient of 0.078. Appendix Table 4 demonstrates that the relationship in this column is robust to numerous other functional forms for the non-linearity.

### 2.3.3 Can Urban Disamenities Explain the Correlation between Unhappiness and Urban Decline?

We next consider whether observable urban disamenities can explain the correlation between unhappiness and urban decline. We add a variety of correlates of decline to our previous specifications and ask whether these variables reduce the coefficient on urban decline. While none of these estimates can be taken as being causal, they represent a rough pass at judging whether the correlation between decline and happiness merely represents the correlation between decline and some other more important variable.

Our first control is January temperature. The correlation between warm weather and metropolitan growth is well known (Glaeser and Tobio 2008), and it is certainly possible that tough winters are depressing. While Oswald and Wu (2010) find that climate has a significant relationship with self-reported happiness, we find no connection in our specification once we have controlled for population growth non-linearly. Moreover, this control does little to the estimated coefficient on population decline.<sup>6</sup> Note that we have a smaller sample in these regressions because we lack crime data for many of our metropolitan areas.

Our second climate variable is precipitation, measured in annual inches of rain. This variable again has little correlation with SWB in our data after controlling for population decline. The third variable we test is the log of the number of serious crimes per capita, and it also has no detectable relationship. While being victimized may certainly make someone unhappy, it seems quite possible that crime is sufficiently concentrated in certain population subgroups that it has little impact on average happiness.

The fourth variable captures pollution, which might well be higher in America's erstwhile industrial heartland. We have tried many difference measures of local pollution levels, but none of them correlate well with happiness. Here we include total particulates (mean of 10 micron particulate matter, from 2000) and it has little correlation with happiness.

Our fifth variable is the Gini coefficient, which measures income inequality as of the year

---

<sup>6</sup>In our working paper (Glaeser, Gottlieb and Ziv 2014) we also show regressions that include these controls one at a time. That approach has no impact on the conclusions we draw here.

2000. While Alesina, Di Tella and MacCulloch (2004) find that happiness decreases with inequality, especially in Europe, we find a slight positive relationship between happiness and inequality across U.S. metropolitan areas. Moreover, controlling for inequality does little to change the estimated impact of population decline on unhappiness. The weak connection between inequality and SWB in the BRFSS is somewhat odd, because it is quite strong in the General Social Survey (Glaeser, Resseger and Tobio 2009).

Column six shows that including all of these variables reduces the coefficient on decline from 0.21 to 0.156. Adding the endogenous demographic controls causes the coefficient to drop further to 0.08. This should be compared with column two's coefficient of 0.134, which is the effect of decline on happiness without these other amenity controls, but with endogenous demographics. Finally, in the last regression, we include state fixed effects, which changes the coefficients on some of the area amenity controls but have little impact on population growth coefficients.

Throughout the specifications in this table, bolstered by the robustness checks in Appendix Table 4, the coefficient on urban decline is statistically robust and the magnitudes remain quite similar. While it is certainly true that controlling for education and family status significantly reduces the estimate coefficient, other individual controls change the coefficient only slightly. We believe that this suggests that this effect is less likely to reflect unobserved heterogeneity, but to address that issue we turn to the NSFH. Urban Decline and Unhappiness with Movers and Stayers

We now ask whether the unhappiness of declining cities appears to be limited to longer term residents or whether they are similar for recent migrants. We explore this question in the first two columns of Table 2.4 using the NSFH. In its first two waves, the NSFH is a clean panel that can, in principle, enable us to look at SWB for people who move between areas. Two significant challenges with the NSFH are that the samples are small and the time between the first and second waves is small (under five years), so the number of movers is smaller still.

The first regression shows the effect of the population growth spline with exogenous

**Table 2.4: Happiness, urban decline, and mobility, NSFH**

| Dependant Variable  | (1)<br>Self-Reported Well-Being | (2)                 | (3)<br>Mover status    | (4)                    | (5)<br>Wave 2 PMSA Growth | (6)                  | (7)                  | (8)                     |
|---|---------------------------------|---------------------|------------------------|------------------------|---------------------------|----------------------|----------------------|-------------------------|
| Change in log population (below median) 1950-2000         | 0.141***<br>(0.0362)            | 0.142**<br>(0.0702) | 0.00942<br>(0.0331)    | 0.00804<br>(0.0333)    | 0.580***<br>(0.176)       | 0.619***<br>(0.178)  |                      |                         |
| Change in log population (above median) 1950-2000         | -0.0574***<br>(0.0160)          | -0.0387<br>(0.0356) | 0.0422***<br>(0.0151)  | 0.0431***<br>(0.0147)  | 0.0139<br>(0.0988)        | -0.00488<br>(0.0997) |                      |                         |
| Mover   |                                 | 0.0869<br>(0.124)   |                        |                        |                           |                      |                      |                         |
| Mover x Change in log population (below median) 1950-2000 |                                 | -0.0851<br>(0.211)  |                        |                        |                           |                      |                      |                         |
| Mover x Change in log population (above median) 1950-2000 |                                 | 0.0217<br>(0.0681)  |                        |                        |                           |                      |                      |                         |
| SWB x Change in log population (below median) 1950-2000   |                                 |                     |                        | -0.0327**<br>(0.0160)  |                           |                      |                      |                         |
| SWB x Change in log population (above median) 1950-2000   |                                 |                     |                        | 0.0152<br>(0.0119)     |                           |                      |                      |                         |
| Wave 1 Subjective Well-being (SWB)                        |                                 |                     | -0.00137<br>(0.00352)  | 0.0128*<br>(0.00723)   | 0.0227<br>(0.0262)        | 0.0218<br>(0.0279)   | 0.00227<br>(0.00150) | 0.00165<br>(0.00149)    |
| Wave 2 Subjective Well-being (SWB)                        | -0.0131***<br>(0.00403)         |                     |                        |                        |                           |                      |                      |                         |
| Wave 1 PMSA BLUP (below median)                           |                                 |                     |                        |                        |                           |                      | 0.170***<br>(0.0598) | 0.152**<br>(0.0583)     |
| Wave 1 PMSA BLUP (above median)                           |                                 |                     |                        |                        |                           |                      | 0.118<br>(0.112)     | 0.108<br>(0.117)        |
| Log Household Income                                      |                                 |                     | -0.0123**<br>(0.00537) | -0.0124**<br>(0.00537) |                           | -0.00476<br>(0.0332) |                      | -0.00353**<br>(0.00171) |
| Additional Controls                                       | No                              | Yes                 | Yes                    | Yes                    | No                        | Yes                  | No                   | Yes                     |
| Observations  | 17,019                          | 8,491               | 8,528                  | 8,528                  | 935                       | 935                  | 1,513                | 1,513                   |
| R-squared   | 0.010                           | 0.046               | 0.066                  | 0.066                  | 0.093                     | 0.113                | 0.039                | 0.062                   |

Source: Authors' regressions on microdata from the National Survey of Families and Households (Sweet and Bumpass, 1996) and U.S. Census (Ruggles et al., 2010).

Notes: All regressions control for month fixed effects, age, race, and sex. "Additional controls" include education, marital status, and family size. Standard errors in parentheses are robust to heteroskedasticity (\*\*\*) p<0.01, \*\* p<0.05, \* p<0.1).

demographic controls. The coefficient is 0.14, somewhat smaller than in the BRFSS, but the question is different and the controls are not identical. This result confirms our baseline finding in a separate dataset.

We next explore whether people who move to a new metropolitan area experience the happiness level of the area where they are newly arrived. To do this, we look at observations in the second wave of the NSFH, where we can distinguish between recent movers and previous residents. In the second regression, we estimate whether the coefficient on decline is different for individuals who moved into the metropolitan area between the first and second wave. The coefficient on decline for stayers is very similar to that in the previous regression. The interaction between the decline measure and being a mover is negative, meaning that decline is less strongly associated with unhappiness for the movers. However, while the interaction is not small, it is not distinct from zero. It is difficult to conclude much

from this regression.<sup>7</sup>

One possible explanation for the relationship between decline and unhappiness is selective migration. Individuals who leave declining cities may be happier than their neighbors, or growing cities may attract individuals who are happier than the population as a whole. The panel nature of the NSFH allows us to test this hypothesis.

Columns three and four use the entire NSFH sample, in order to test whether individual and PMSA characteristics in wave 1 predict whether an individual moves between waves 1 and 2. Both columns control for subjective well-being in wave 1, the population growth spline of an individual's wave 1 PMSA, our standard set of individual controls, and income in wave 1. In column three, the upper part of the population growth spline positively predicts mover status, reflecting the high degree of population churn in the upper tail of growing cities. Subjective well-being in wave 1 is not predictive of whether an individual will move between waves 1 and 2.

Column four adds an interaction between individual wave 1 subjective well-being and the population growth spline. Subjective well-being is now marginally positive and significant. Critically, the interaction between the lower spline and subjective well-being is negative and significant. Happier people are less likely to leave declining cities, relative to rising cities. Put another way, we can reject the hypothesis that the happiest individuals are selectively moving out of declining areas.

In the remaining columns, we focus on the subsample of the NSFH who moved MSAs between waves 1 and 2. In columns five and six, we test for the hypothesis that happier migrants select growing cities. We focus on the 935 movers in our sample for which we have data on PMSA population for waves 1 and 2. In column five, we use our set of exogenous controls, wave 1 individual subjective well-being, and wave 1 PMSA population growth spline. In column six, we add controls for wave 1 endogenous individual characteristics and income. Although we cannot reject a positive relationship between wave 1 subjective

---

<sup>7</sup>In Table 5 of the working paper version (Glaeser Gottlieb and Ziv 2014), we run additional robustness checks on these regressions, including additional controls, fixed effects, and sample restrictions. These have little impact on our conclusion here.



well-being and wave 2 PMSA population growth, we find no evidence for selection of individuals with higher subjective well-being into growing cities.

Finally, columns seven and eight analogously assess whether happier migrants select happier cities using data from all 1513 movers in the NSFH. In this specification, we also find little connection between wave 1 “happiness” for movers and choosing, conditional upon moving, to relocate to a happier locale. In Column eight, we add controls for endogenous characteristics and income in wave 1. The positive relationship between wave 1 and wave 2 subjective well-being decreases in size and continues to be insignificant. The data do not support the hypothesis that unhappy migrants choose declining or unhappy cities, but the results are not strong enough to reject the possibility of selective migration.

Table 2.5 now turns to a different data set, the General Social Survey (GSS). The public version of the GSS contains state name and city level population. These two variables enable us to predict the population decline in the area with a fairly high degree of accuracy for the overwhelming majority of data points.

**Table 2.5: Happiness regressions using the General Social Survey**

|  | (1)                     | (2)                  | (3)                  | (4)                   | (5)                   |
|--|-------------------------|----------------------|----------------------|-----------------------|-----------------------|
| Dependent Variable   | Self-Reported Happiness |                      |                      |                       |                       |
| Change in log population (below median) 1950-2000                        | 0.214***<br>(0.0527)    | 0.222***<br>(0.0818) | 0.521***<br>(0.104)  | 0.485***<br>(0.113)   | 0.459***<br>(0.113)   |
| Change in log population (above median) 1950-2000                        | -0.0295<br>(0.0438)     | -0.0382<br>(0.0658)  | -0.0961<br>(0.0768)  | -0.0819<br>(0.0853)   | -0.0752<br>(0.0803)   |
| Change in log population (below median) 1950-2000 *<br>Individual Moved  |                         | -0.0355<br>(0.118)   |                      |                       |                       |
| Change in log population (above median) 1950-2000 *<br>Individual Moved  |                         | 0.0543<br>(0.0780)   |                      |                       |                       |
| Moved  |                         | 0.0378<br>(0.0555)   |                      |                       |                       |
| Change in log population (below median) 1950-2000 * 1980<br>Decade Dummy |                         |                      | -0.287**<br>(0.134)  | -0.296**<br>(0.123)   | -0.277**<br>(0.122)   |
| Change in log population (below median) 1950-2000 * 1990<br>Decade Dummy |                         |                      | -0.372***<br>(0.121) | -0.451***<br>(0.0992) | -0.421***<br>(0.0952) |
| Change in log population (below median) 1950-2000 * 2000<br>Decade Dummy |                         |                      | -0.556***<br>(0.144) | -0.446**<br>(0.169)   | -0.440***<br>(0.159)  |
| Change in log population (below median) 1950-2000 * 2010<br>Decade Dummy |                         |                      | -0.510***<br>(0.172) | -0.344<br>(0.225)     | -0.298<br>(0.214)     |
| Change in log population (above median) 1950-2000 * 1980<br>Decade Dummy |                         |                      | 0.130<br>(0.103)     | 0.134<br>(0.103)      | 0.128<br>(0.0969)     |
| Change in log population (above median) 1950-2000 * 1990<br>Decade Dummy |                         |                      | 0.0398<br>(0.0978)   | 0.113<br>(0.128)      | 0.111<br>(0.118)      |
| Change in log population (above median) 1950-2000 * 2000<br>Decade Dummy |                         |                      | 0.0107<br>(0.0795)   | 0.0116<br>(0.110)     | 0.00589<br>(0.106)    |
| Change in log population (above median) 1950-2000 * 2010<br>Decade Dummy |                         |                      | 0.188**              | 0.161                 | 0.180                 |
| Additional Controls  | No                      | Yes                  | No                   | Yes                   | Yes                   |
| Income and Employment Controls   | No                      | Yes                  | No                   | No                    | Yes                   |
| Observations   | 9,995                   | 7,541                | 9,995                | 7,541                 | 7,541                 |
| R-squared  | 0.021                   | 0.051                | 0.024                | 0.040                 | 0.054                 |

Source: Authors' regressions on microdata from the General Social Survey (GSS) and U.S. Census (Ruggles et al., 2010).

Notes: All regressions control for a year trend, age, race, and sex. "Additional controls" include education, marital status, and family size. Standard errors in parentheses are clustered at the MSA level (\*\*\* p<0.01, \*\* p<0.05, \* p<0.1).

In regression one, we again estimate the spline controlling for exogenous individual attributes, this time in the GSS. We continue to find a strong positive relationship between growth and happiness for the areas where growth rates are below the sample median. The second regression interacts these variables with an indicator variable denoting whether the individual has moved metropolitan areas since age 16. The interaction between this variable

and population decline is negative, but very close to zero. In this larger sample, we see little evidence suggesting that the unhappiness associated with urban decline is limited to longer term residents.

#### **2.3.4 Is the Unhappiness of Declining Cities New or Old?**

Historical data can help us assess whether the relationship between unhappiness and decline reflects the impact of decline itself, or whether these now declining cities were historically defined more by productivity than by pleasure. According to the first view, Detroit was once a place of happiness as well as prosperity, but as the prosperity declined and the social problems increased, unhappiness spread. According to the second view, Detroit was unhappy even during its heyday, but historically, its residents were well compensated for their joylessness. Capital and labor located in the city historically because it had natural advantages, such as access to waterways, that made up for the loss in happiness.

The era of comprehensive urban happiness measures really only began ten years ago with the BRFSS. The NSFH goes back 20 years, but even that is a relatively short historical window. To investigate the more distant past, we turn to the General Social Survey (GSS). The GSS has relatively comprehensive personal controls, but still dates back only to the early 1970s.

Our approach is to estimate the impact of area-level population change and then to examine how this effect changes over the decades. We do this using the General Social Survey in the last three regressions of Table 2.5. We again estimate a spline for population growth, but we interact the coefficients on that spline with indicator variables that represent each decade. The population growth is defined over the entire 1950-2000 period, but the interactions allow the connection between decline and happiness to differ across the decades.

In Regression 3, we control for standard demographic variables and a year trend variable. As the regression shows, the interaction is strongly negative after the 1970s, meaning that the correlation between unhappiness and decline has decreased over time. Indeed, by the 2000s, the connection has disappeared entirely, which is of course, not what we observed

in the BRFSS. This shows that the cities that are declining over the entire period were unhappier in the 1970s, relative to other areas, than they were after 2000. These results are compatible with the view that unhappiness caused the decline or that declining cities have long-standing attributes associated with unhappiness, but they seem don't seem compatible with the view that unhappiness has grown following decades of decline.

Regression four includes controls for the endogenous demographics. While the overall negative relationship weakens, the time pattern is unchanged. The final regression includes controls for income and unemployment. Again, the basic time pattern remains clear. Declining cities were even unhappier in the past than they are today.

Our working paper presents results from even farther in the past. Using Gallup surveys from the 1940s, we show a significant negative connection between unhappiness and city population during those years, although that connection is not stronger in states that experienced more subsequent decline (Glaeser, Gottlieb and Ziv 2014). Nonetheless, eight of the ten largest U.S. cities in 1950 lost at least one-fourth of their population over the next 50 years. So these results support the view that the large cities of the 1940s, which typically experienced subsequent decline, were also places marked by somewhat lower happiness levels during their heyday.

These results are hardly definitive, but taken together they suggest that urban unhappiness is not exclusively recent. The GSS shows larger results in the past than in the present. This corresponds to results we can see in the BRFSS estimates. The correlation between log metropolitan area population in 2010 and adjusted life satisfaction is 0.03. The correlation between that same happiness outcome and log area population in 1950 is -0.28.

## **2.4 Why Does Happiness Differ Across Space?**

If self-reported happiness has any equivalence to the economist's concept of utility, then modestly enduring differences in self-reported life satisfaction seem to challenge the view that migration and the free operation of housing markets ensure that utility levels are equalized across space. Alternatively, if there are persistent differences in subjective well-

being for identical people across space and a spatial equilibrium does hold, then this would imply that subjective well-being is just not equivalent to the economists' conception of utility.

Perhaps the differences that we measured above may not really represent differences in subjective well-being among otherwise identical human beings. The residents of declining cities may have less marketable skills, of various forms, than residents of growing cities, and as such, they would naturally earn less and have lower levels of life-satisfaction or utility in any metropolitan area. Yet our results control for a bevy of individual characteristics and controlling for added metropolitan area level variables, including the percent with college degrees or the share of the population that is white, only modestly reduces the relationship between decline and self-reported life satisfaction. The estimated relationship actually increases in magnitude if we restrict our samples to metropolitan areas with relatively similar levels of college graduates.<sup>8</sup> Finally, while the individual fixed effects results on urban decline were inconclusive outside of the larger cities, there are still significant differences in SWB across cities when we control for individual fixed effects.

It is also possible that individuals on the margin of moving across areas receive the same welfare, but that infra-marginal individuals differ in their average level of well-being across space. Yet for this view to be correct, we would need an explanation of why the average infra-marginal welfare in declining areas is significantly lower than in growing areas, even if the marginal happiness levels are the same.

Another interpretation is that when equivalent individuals made location decisions, their expected happiness was equal across space, but ex post some migrants have fared worse than others, either because they were bad at projecting the happiness that different places bring, or because some areas have received particularly adverse shocks. According to this view, ex post welfare differs across space, even though ex ante welfare does not. But if this

---

<sup>8</sup>The coefficient when our happiness variable is regressed on the change in population between 1950 and 2000 is 0.023. When we control for share of the population with college degrees and percent white, the coefficient drops to 0.02. When we restrict our sample to metropolitan areas in which twenty-to-thirty percent of adults have college degrees, then the coefficient rises to 0.031.

view is correct, then the connection between urban decline and unhappiness should exist primarily for longer term residents of the area, such as people who are unlikely enough to be born in metropolitan areas in decline, not recent migrants. Yet as we have discussed, the connection between unhappiness and urban decline is stronger for individuals who chose to come to the area as adults than for individuals who were born into the area. Moreover, given the well-advertised urban problems of many declining cities such as Detroit, it is hard to imagine that migrants are all that surprised—although it is certainly true that there are general problems in forecasting happiness (Gilbert 2006; Kahneman and Krueger 2006).

#### **2.4.1 Is Happiness a Measure of Utility?**

While we acknowledge that the preceding facts are open to multiple interpretations, we focus on one in particular: individuals maximize neither happiness nor life satisfaction, and will sacrifice both for the right reward. In this view, individuals in less happy areas forego well-being in exchange for some other advantage. While we do not suggest that SWB and happiness are meaningless concepts, this view posits that their meaning is distinct from economists' conception of utility—which is merely a representation of choice. In standard microeconomic theory, an outcome yields higher utility if and only if it is preferred.

The debate over whether individuals either do or should maximize happiness is ancient and intellectually rich. Bentham (1789) famously wrote that “It is for [pain and pleasure] alone to point out what we ought to do, as well as to determine what we shall do.” Bentham's claim that happiness is both the positive and normative determinant of behavior greatly influenced 19th century economists such as John Stuart Mill.

These economists reflect an ancient philosophical tradition, dating back to the Cyrenaic School. The Cyrenaics emphasized pleasure as life's central goal and influenced the later Epicureans. Giants of medieval philosophy including Augustine and Aquinas accepted that human beings pursued happiness above all, but taught that true happiness is to be found by following God's will.

Some modern researchers on happiness, though certainly not all, have also conflated

happiness with utility (Alesina, Di Tella and MacCulloch 2004) or at least social welfare (Easterlin 1995).<sup>9</sup> Yet it is quite possible to believe that happiness is interesting and important, without accepting the equivalence. There is also an equally ancient and distinguished philosophical tradition rejecting the notion that individuals either do or should maximize happiness.

While the Epicureans believed in maximizing pleasure and minimizing pain, the Stoics did not. About 1900 years ago, Epictetus wrote “What is our nature? To be free, noble, self-respecting. . . . We must subordinate pleasure to these principles. . . .” Epictetus is making the normative claim that other goals—freedom, nobility, and self-respect—trump happiness.

Economists from Fisher (1892) through Stigler (1950) and Becker and Rayo (2008) have followed this approach. Becker and Rayo wrote of “an alternative interpretation of the happiness data, namely, that happiness is a commodity in the utility function in the same way that owning a car and being healthy are.” Perhaps the best known evidence supporting this interpretation is that parents of small children typically report lower happiness or life satisfaction (Baumeister 1991).<sup>10</sup> If happiness measured utility, then presumably this relationship should impede the survival of the species. Yet in Becker and Rayo’s formulation, this negative relationship is no puzzle at all—parents receive ample compensation, in the form of progeny, for their suffering.

Leaving aside survey evidence on SWB, we can look for reasonably direct evidence of low utility levels by measuring suicide (Becker and Posner 2004). If suicide reflects low

---

<sup>9</sup>Alesina et al. (2014, p. 2010) explicitly state that they “measure ‘utility’ in terms of survey answers about ‘happiness’.” They elaborate in footnote 7 that they, and in their view much of the literature on economics of happiness, aim to measure “experienced utility, a concept that emphasizes the pleasures derived from consumption”. They view these survey responses, in certain circumstances, as “reasonable substitutes to observing individual choices” (footnote 7, p. 2012). Many other prominent papers in this literature implicitly posit such an equivalence, such as Easterlin (1995, p. 36). Easterlin writes, “Formally, this model corresponds to a model of interdependent preferences in which each individual’s utility or subjective well-being varies directly with his or her own income and inversely with the average income of others.” Of course the literature has also considered many subtle points about the appropriate conception of subjective well-being. For example, Kahneman and Thaler (1991) and Kahneman and Krueger (2006) distinguish between decision utility and experienced utility. We certainly do not claim to introduce a novel distinction here. Our contribution, in part, is to use the decision-utility maximization embodied in spatial equilibrium to put more structure on the theoretical and empirical relationships between choices and subjective well-being.

<sup>10</sup>But see Deaton and Stone (2014) on the importance of sorting into parenthood.

utility, then the relationship between suicide and self-reported happiness provides another way to consider whether low SWB captures low utility. In the working paper version (Glaeser, Gottlieb and Ziv 2014), we show that suicide rates are uncorrelated with subjective well-being across metropolitan areas, corroborating Daly et al.'s (2011) cross-state evidence.

Some suggest that in its very wording, life satisfaction should capture all the elements of utility. While it seems implausible to hope that maximizing utility should automatically mean maximizing joy or happiness, might individuals answer the question about life satisfaction in such a way that actually ranks their preferred outcomes, as does a utility function? If so, an individual who has received a preferred outcome will report a higher level of life satisfaction. Hence utility and subjective well-being converge because well-being acts as a barometer to measure how well people have achieved their goals.

Yet this view seems barely more tenable than the view that happiness should miraculously map onto human preferences. Someone may choose a more competitive environment with more opportunity to shape the world, and yet know that this environment will—by opening up opportunities and inviting comparisons with high achievers—lead to less satisfaction. A rational person could select a Ph.D. program, or a city, despite recognizing that it will lead to less satisfaction.

Among members of the classical economic tradition, Bernard de Mandeville may be the most powerful proponent of the view that human beings should not maximize happiness, especially not in location choice. In *The Fable of the Bees* he writes, “To be happy is to be pleas’d, and the less Notion a Man has of a better way of Living, the more content he’ll be with his own... the greater a Man’s Knowledge and Experience is in the World, the more exquisite the Delicacy of his Taste, and the more consummate Judge he is of things in general, certainly the more difficult it will be to please him... But when a Man enjoys himself, Laughs and Sings, and in his Gesture and Behaviour shews me all the tokens of Content and Satisfaction, I pronounce him happy, and have nothing to do with his Wit or Capacity.” Clearly, de Mandeville thinks little of happiness. When he writes “ask’d where I thought it was most probable that Men might enjoy true Happiness, I would prefer a small



peaceable Society, in which Men, neither envy'd nor esteem'd by Neighbours, should be contented to live upon the Natural Product of the Spot they inhabit, to a vast Multitude abounding in Wealth and Power," he is not espousing such places, but arguing that it is perfectly sensible to choose busier, but less happy, locales.

We don't dispute the desirability of happiness, and for that reason, the spatial equilibrium logic of Rosen (1979) and Roback (1982) implies that there must be some compensation offsetting the unhappiness of declining cities. Residents must receive some other benefit, such as higher real wages, that offsets the costs of lower life-satisfaction. Otherwise, it would be hard to understand why they remain in unhappy cities. We formalize these issues in the model that follows.

### 2.4.2 Happiness and Utility

To formalize this discussion, we begin with a general framework meant to capture the difference between happiness and utility. We then adapt our structure to deal with cities and urban decline, which requires considerably more assumptions about structure and ultimately even functional forms. This latter section puts forward the model that will be taken to the data in Section IV.

In Becker (1965), individuals maximize a function  $U(\cdot)$  defined over a vector of objectives  $\tilde{Z}$ , where each element in that vector  $Z_i$  is a function of time ( $t_i$ ) and spending ( $s_i$ ). One possible approach is to assume that life satisfaction is defined over an alternative function  $H(\cdot)$  of those same objectives, but that approach provides little guidance for modeling or testable implications.

We assume that subjective well-being represents an alternative function  $W(\cdot)$  over the same set of objectives. It may be that welfare is a function of well-being and other objectives, or that well-being is simply a slightly different function of exactly the same inputs that guide utility. In the first case, utility can be described as  $U(W(\tilde{Z}), \tilde{Z}_{NH})$ , where  $\tilde{Z}_{NH}$  refers to objectives that enter into utility directly, such as child-rearing, as well as possibly also impacting well-being.

We will approach well-being and utility as reflecting a combination of experiences and achievements. Well-being or happiness will be conceived as experience-based utility, following Bentham (1789) and Kahneman and Krueger (2006). Individuals care about experienced utility, but they also care about achievements, which can also be produced with time and money. We lose little generality by assuming at this point that there is a single achievement, which is produced with achievement-specific time denoted  $t_A$  and achievement-specific spending  $s_A$ . Individual earnings are the product of wages and time spent working  $wt_w$  and unearned income  $y_0$ , which includes the fixed cost of housing.

Time spent working and time pursuing the alternative achievement convey experienced utility of  $t_w$  and  $t_A$  per unit time. The remainder of hedonic time generates well-being equal to  $h(s_h)t_h$ , where  $s_h$  reflects the total amount of spending on these activities. The term  $h(s_h)t_h$  is meant to aggregate all other time, including sleeping. So the individual's problem is to maximize:

$$U(h_w t_w + h_A t_A + h(s_h) t_h, Z(s_a, t_a)) \quad (2.5)$$

subject to the time budget constraint  $t_w + t_a + t_h = 1$  (where total time available is normalized to one) and the cash budget constraint  $wt_w + y_0 = s_h + s_a$ . The two budget constraints can be combined to create a single total budget constraint of  $w + y_0 = w(t_a + t_h) + s_a + s_h$ .

In this model, as in almost all economic models, more income is preferred to less, and translates into higher levels of utility. Yet the link between happiness and wages is less clear. If, for example,  $Z(\cdot)$  is produced entirely with earnings, then as long as the uncompensated wage elasticity is positive, happiness diminishes with wages even though utility increases. If  $h(s_h) = h_0$  is independent of income, then the derivative of happiness with respect to the wage equals  $(h_0 - h_w)$  times the derivative of  $t_h$  with respect to the wage, which equals

$$\frac{\partial t_h}{\partial w} = \frac{-wZ'(s_A)U_Z + (1-t_h)Z'(s_A)((h_0-h_w)U_{HZ} - wZ'(s_A)U_{ZZ})}{-(U_{HH}(h_0-h_w)^2 - 2(h_0-h_w)wZ'(s_A)U_{HZ} + (wZ'(s_A))^2U_{ZZ})} \quad (2.6)$$

Across space, the impact of income on happiness may be even more negative. Suppose that amenities are constant across space, and that utility levels are unchanged with changes in wages;  $\frac{\partial y_0}{\partial w} = -(1-t_0)$ : the change in housing costs exactly offsets the change in earnings. If this is the case, then in the case discussed above, where spending does not impact the hedonic flow of time,  $\frac{\partial t_h}{\partial w} = \frac{-wZ'(s_A)U_Z}{-(U_{HH}(h_0-h_w)^2 - 2(h_0-h_w)wZ'(s_A)U_{HZ} + (wZ'(s_A))^2U_{ZZ})} < 0$ , so happiness is always lower in higher wage cities. Since the impact of area level wages is compensated, rather than an uncompensated change in wages, it will invariably cause an increase in hours worked and a decrease in time spent in household production.

We now turn to the spatial equilibrium, where we assume that  $h_w = h_a = 0$ . We also assume a Cobb-Douglas utility function, with a weight of  $\alpha$  on happiness, power functions for producing the other goods, and that  $Z(s_a, t_A) = z_0(s_A)^z(t_A)^{1-z}$ , where  $z_0$  is a city-specific production shifter. We assume that time spent at work is fixed at  $\hat{t}_w$  but that time can still be allocated between leisure and the other achievement. Further,  $h(w\hat{t}_w + y_0) = h_0(w\hat{t}_w + y_0)^\gamma$ , where  $h_0$  is a city-specific amenity. Given these assumptions, then indirect utility is proportional to  $(h_0)^\alpha(z_0)^{1-\alpha}(w\hat{t}_w + y_0)^{\alpha\gamma+(1-\alpha)z}$  and happiness is proportional to  $h_0(w\hat{t}_w + y_0)^\gamma$ .

The Cobb-Douglas welfare function generates a happiness-income tradeoff of  $\frac{d\log(\text{Happiness})}{d\log(\text{Income})} = -\frac{1-\alpha}{\alpha}$ . This tradeoff is a distinct concept from the derivative of happiness with respect to the wage (assuming unearned income is negligible), which equals  $\gamma$ .

We have two options here, choosing fixed or flexible working time, but the simpler functional forms come with fixed hours. In that case, the spatial equilibrium condition can be written as:

$$w\hat{t}_w + y_0 = k_0 h_0^{-\frac{\alpha}{\alpha\gamma+(1-\alpha)z}} z_0^{-\frac{1-\alpha}{\alpha\gamma+(1-\alpha)z}}. \quad (2.7)$$

The values of  $h_0$  and  $z_0$  are determined both by natural amenities, such as climate, and amenities tied to public services, such as safety. Declining areas could well have lower levels of quality of life both because they are in relatively cold areas of the U.S. and because a reduced level of spending leads to lower levels of public amenities.

An urban equilibrium involves three separate equations. The first is the spatial equilibrium curve for consumers in which welfare—but not happiness—must equal a constant reservation utility across space. The second condition is that firm profits are equalized across space. The third condition is that the cost of housing equals the cost of supplying homes.

We assume a linear housing supply curve, so that the flow cost of housing in a city, denoted  $r$ , is  $r = c_0 + c_1 \log(N_t) + c_2 \log(N_t/N_{(t-1)})$ , where  $N_t$  reflects the population in the place, and we assume that  $y_0 = -r$ . This can be generated by an assumption that houses are created with a Cobb-Douglas utility function using traded and non-traded capital, where non-traded capital is in fixed supply. In principle,  $c_0$ ,  $c_1$  and  $c_2$  might all vary across areas.

Finally, we have linear labor demand so that  $w\hat{t}_w = A - B \log(N_t)$ . This can be generated by assuming that there are a fixed number of firms with Cobb-Douglas production functions and two types of labor, one of which is traded and the other is not (Glaeser 2007). Again,  $A$  and  $B$  might differ across metropolitan areas.

Using the housing supply curve, labor demand curve, and taking logs of equation (7) yields:

$$\log(N_t) = \frac{1}{B + c_1 + c_2} (A - c_0 + c_2 \log(N_{(t-1)}) - k_0 h_0^{-\frac{\alpha}{\alpha\gamma + (1-\alpha)z}} z_0^{-\frac{1-\alpha}{\alpha\gamma + (1-\alpha)z}}) \quad (2.8)$$

$$w\hat{t}_w = \frac{1}{B + c_1 + c_2} ((c_1 + c_2)A + Bc_0 - Bc_2 \log(N_{(t-1)}) + Bk_0 h_0^{-\frac{\alpha\gamma}{\alpha\gamma + (1-\alpha)z}} z_0^{-\frac{1-\alpha}{\alpha\gamma + (1-\alpha)z}}) \quad (2.9)$$

$$r = \frac{1}{B + c_1 + c_2} ((c_1 + c_2)(A - k_0 h_0^{-\frac{\alpha}{\alpha\gamma + (1-\alpha)z}} z_0^{-\frac{1-\alpha}{\alpha\gamma + (1-\alpha)z}} + B(c_0 - c_2 \log(N_{(t-1)}))) \quad (2.10)$$

$$\log(Happiness) = \log(k_0^\gamma(1 - \hat{t}_w)) + \frac{(1 - \alpha)z}{\alpha\gamma + (1 - \alpha)z} \log(h_0) - \frac{(1 - \alpha)\gamma}{\alpha\gamma + (1 - \alpha)z} \log(z_0). \quad (2.11)$$

Population is increasing with productivity, decreasing with the cost of providing housing and increasing with the two amenity variables. Income is rising with productivity and housing cost, and falling with amenities. Housing rents increase with productivity, with the cost of housing and with amenities. Happiness is rising with the happiness-related amenity and declining with the non-happiness-related amenity. This becomes a four-equation system for empirical work, where the impacts of local variables can be traced through these four distinct outcomes.

In this formulation, happiness is a measure of local amenities—and local amenities only—because population and housing prices adjust to shifts in local demand and construction costs. The spatial equilibrium requires that gaps in real income end up being proportional to happiness, holding as such happiness should be declining in real income. The slope is predicted to equal  $\frac{(1-\alpha)z}{\alpha}$  on real income, which equals  $\frac{(1-\alpha)}{\alpha}$ —the basic happiness-income tradeoff—times  $z$ —the elasticity of the non-happiness-related component of welfare with respect to earnings.

We can also use the spatial indifference condition and find that happiness is proportional to  $(z_0)^{(\alpha-1)/\alpha} (w\hat{t}_w + y_0)^{(\alpha-1)/\alpha z}$ . Holding  $z_0$  constant, we expect to find that richer places are less happy, and holding income constant, we expect to find happier places deficient in some other desirable (non-happiness related) amenity. The unhappiness of declining cities, therefore, needs to be compensated either with higher real incomes or with some other asset.

## 2.5 Are Individuals Compensated for Unhappiness?

In this section, we do not implement the four equation empirical estimation section discussed above, but rather restrict ourselves to a simpler empirical approach inspired by the theory. We will test whether individuals in declining or unhappy cities are being compensated for

their misery by either lower housing costs or higher wages.

This approach begins with a simple view of America's changing urban system. We initially built cities in places where firms had a productive advantage because of proximity to waterways and coalmines. Moreover, we also built those cities in ways that favored productivity rather than pleasure. Over time, declining transport costs enabled capital and labor to flee low amenity places (Glaeser and Kohlhase 2004) and move to "consumer cities" endowed with higher amenity levels (Glaeser, Kolko and Saiz 2001). An increasingly wealthy population also built new cities that were more oriented towards consumer well-being.

Within the context of the model, this can be understood as a change in the covariance between productivity and the amenity parameters. In early 20th century America, productivity may have been higher in lower amenity places but in late 20th century America, that negative covariance disappeared. As a result, population growth was faster in places that had higher amenities initially and lower levels of productivity.

This argument provides a slightly different interpretation of Easterlin (1974), at least insofar as it applies to America's metropolitan areas. In the early 20th century, a city needed to be unpleasant to be productive. In the late 20th century, it did not. Since technological change favored pleasant, happier locales, it seemed as if happiness was tied to income growth, even if it was ultimately driven by the local environment.

### **2.5.1 Results from Historical and Contemporary Census Data**

We now turn to the question of compensation. Since no one would presumably have built an intrinsically unhappy city unless it was more productive, we first look at income in 1940. We test whether declining cities, which seem also to have been unhappy in the past, paid higher wages in 1940. Columns 1 and 2 of Table 2.6 show these results (which we expand on in Glaeser, Gottlieb and Ziv 2014). In both columns, we look at earnings for males aged 25 to 55. We include a full battery of controls for age, race and education. Column 1 shows that as population growth increases by 0.1 log point, for cities with population growth below the median, wages drop by 0.014 log points. In 1940, the cities that would subsequently decline

were very well paid. We do not mean to imply causality with this regression. The wage outcomes precede the decline and may have caused the decline. We only mean to suggest that residents of cities that declined after 1950, and that are unhappy today, were relatively well compensated in 1940.

**Table 2.6:** *Income, housing costs, population growth, and happiness, 1940 to 2000 Censuses*

|  | (1)                          | (2)                   | (3)                  | (4)                  | (5)                          | (6)                 | (7)                  | (8)                   | (9)                 | (10)                 |
|--|------------------------------|-----------------------|----------------------|----------------------|------------------------------|---------------------|----------------------|-----------------------|---------------------|----------------------|
|  | Regressions from 1940 Census |                       |                      |                      | Regressions from 2000 Census |                     |                      |                       |                     |                      |
| Dependent variable:                                | Log Income                   |                       | Log House Value      | Log Housing Rent     | Log Income                   | Log House Value     | Log Housing Rent     | Log Income            | Log House Value     | Log Housing Rent     |
| Population Growth 1950-2000, Below Median          | -0.144**<br>(0.0577)         | -0.104***<br>(0.0324) | -0.219*<br>(0.128)   | -0.294***<br>(0.103) | 0.0109<br>(0.0533)           | 0.302**<br>(0.132)  | 0.208**<br>(0.0807)  |                       |                     |                      |
| Population Growth 1950-2000, Above Median          | -0.0442<br>(0.0455)          | -0.0261<br>(0.0404)   | -0.157<br>(0.124)    | -0.0619<br>(0.119)   | 0.00640<br>(0.0196)          | 0.0953<br>(0.0579)  | 0.171***<br>(0.0467) |                       |                     |                      |
| Indicator for MSA with population $\geq 5$ million |                              | 0.0671***<br>(0.0211) | 0.296***<br>(0.0656) | 0.376***<br>(0.0541) | 0.0900***<br>(0.0310)        | 0.555***<br>(0.123) | 0.387***<br>(0.0514) | 0.0950***<br>(0.0284) | 0.489***<br>(0.134) | 0.338***<br>(0.0617) |
| MSA adjusted life satisfaction                     |                              |                       |                      |                      |                              |                     |                      | 0.154<br>(0.197)      | 0.157<br>(0.387)    | 0.659*<br>(0.335)    |
| Observations (thousands)                           | 7,766                        | 7,766                 | 6,535                | 10,527               | 30,044                       | 42,914              | 22,125               | 30,044                | 42,914              | 22,125               |
| R <sup>2</sup>                                     | 0.211                        | 0.215                 | 0.035                | 0.095                | 0.217                        | 0.411               | 0.108                | 0.217                 | 0.400               | 0.089                |

Source: Authors' regressions on U.S. Census microdata (Ruggles et al. 2010).

Notes: Rent variable is monthly contract rent. The 1940 Census data do not include housing quality information. The sample for income regressions is employed males aged 25-55 who work full-time and earn more than half the federal minimum wage for a full-time worker. All income regressions include age, race, and education controls. All regressions include household or person weights. Standard errors are clustered at the MSA level (\*\*\*)  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

The second regression repeats this exercise controlling (coarsely) for city size today. In this column, we simply include an indicator variable that takes on a value of one if the population of the metropolitan area is greater than five million. The coefficient on population is quite large, and it causes the coefficient on population growth below the median to decline to -0.1. The strong positive coefficient on large population size is also quite compatible with the compensation hypothesis, for the Gallup data show that people who live in extremely large cities were dramatically less happy in the 1940s (Glaeser, Gottlieb and Ziv 2014).

We next look at housing costs in that year to test whether either wages or rents compensate the residents in unhappy declining cities. Regressions 3 and 4 show these results. With both measures, cities with subsequent decline had higher housing costs in 1940. These higher costs would mean lower real incomes in these areas, which should eat away some of the compensation received for living in less happy places. However, we are wary of putting

much weight on these regressions since we are not able to effectively control for housing quality in the 1940 Census data. As such, the results could just be showing that the residents of industrial cities in 1940 had substantially better housing than the residents of the Sunbelt in those years.

Column 5 shows the result for income in the 2000 Census microdata. Declining metropolitan areas are not particularly well or poorly paid relative to the U.S. overall. Columns 6 and 7 show results for housing costs. The house value regression estimates a coefficient of 0.3 on population growth below the median, implying that a 0.1 log decrease in population growth between 1950 and 2000 is associated with housing values that are 0.03 log points lower. The result is similar in column 7 when using rents. The residents of declining cities may be less happy, but they are being at least modestly compensated for lower levels of happiness with lower housing costs.

The last three regressions in Table 2.6 look at whether rents or incomes seem to be directly compensating for happiness in 2000. We now correlate the Census outcomes with area happiness itself. We have again controlled for individual attributes, but distinct issues with this specification remain. If the spatial equilibrium is imperfect, then a temporary shock to local income should make people happier—not unhappy. As such, we may not see the negative relationship between happiness and real incomes that is predicted by the model.

Column 8 shows that the happiness variable is not particularly correlated with area income.<sup>11</sup> One interpretation of this is that while people need higher real incomes to put up with unhappy places, these higher incomes also make them happier. These two offsetting forces could create the near-zero coefficient. The remaining regressions examine housing costs, and the results are positive although estimated imprecisely. The coefficient relating rents to happiness in column 10 is 0.66. A 0.1 increase in area happiness is associated with a 0.07 log point increase in rents.

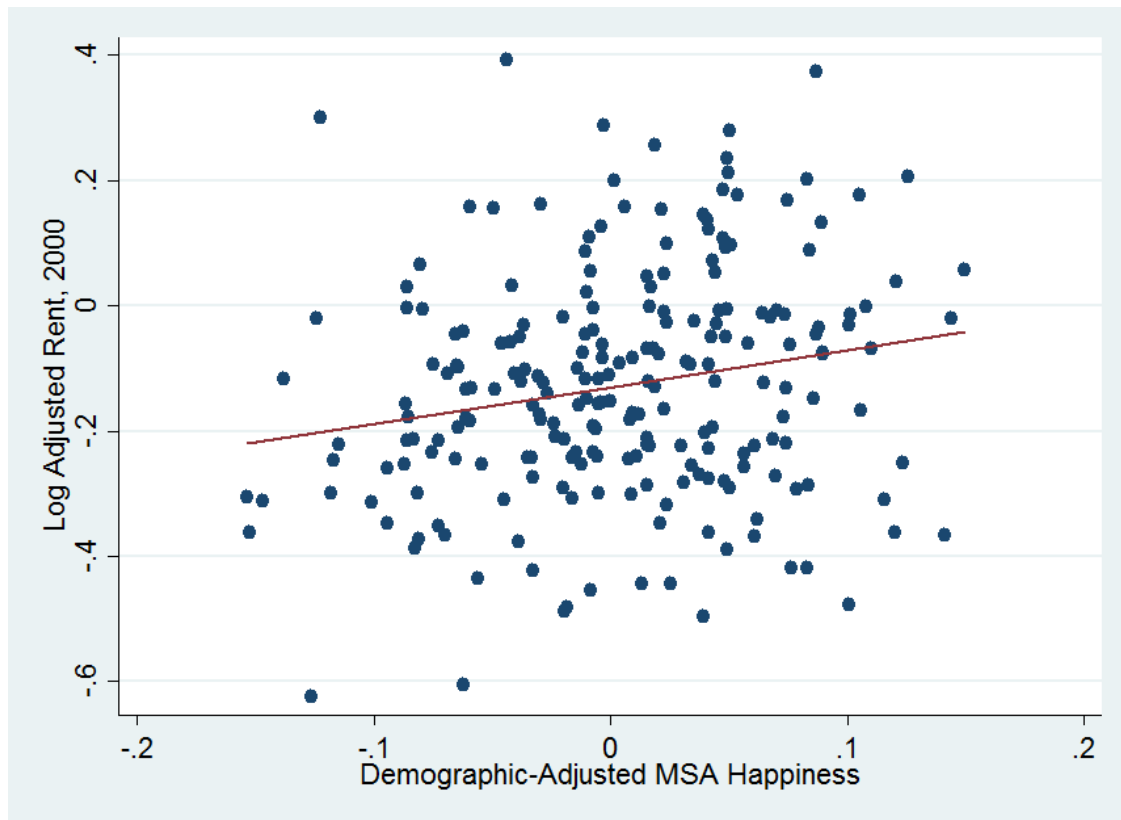
This relationship is shown in Figure 2.4 which collapses the data to the metropolitan

---

<sup>11</sup>Glaeser and Gottlieb (2008) show that happiness and income are also uncorrelated across cities in the GSS.



area level, but excludes California (which has much higher rents). The positive relationship is generally visible, but a small number of cities with higher rents and lower levels of self-reported happiness can be seen in the upper left hand corner of the graph. These are the large metropolitan areas of the east coast, such as New York and Boston. These places also tend to pay high wages, which is presumably how their residents are compensated for lower levels of happiness.



**Figure 2.4:** *MSA rent vs adjusted happiness*

In sum, this table suggests that higher wages compensated for the unhappiness of cities that were large and productive in the 1940s, but would subsequently decline. The population decline has not offset the unhappiness, but is associated with lower housing costs that could partly compensate for the lower reported well-being in such places. This tradeoff is consistent with a model in which happiness is one argument in utility, but harder

to reconcile with views that emphasize happiness as equivalent to utility, or as individuals' ultimate objective.

## 2.6 Conclusion

In this paper, we have documented significant differences in self-reported well-being across American cities that persist when we control for individual demographics and even for individual fixed effects. These facts are not reliably correlated with many area level attributes, but they do seem to be connected with urban decline across at least three large data sets. We do not interpret this correlation as evidence that population decline causes unhappiness. Indeed, cities that have declined also seem to have been unhappy in the past, which suggests that a better interpretation might be that these areas were always unhappy—and that was one reason why they declined.

Differences in happiness or subjective well-being across space weakly support the view that these desires do not uniquely drive human ambitions. If we chose only that which maximized our happiness, then individuals would presumably move to happier places until the point where rising rents and congestion eliminated the joys of that locale. An alternative view is that humans are quite understandably willing to sacrifice happiness or life satisfaction if the price is right. This view rationalizes the well-known tendency of parents to report lower levels of happiness and life satisfaction. Indeed, the residents of unhappier metropolitan areas today receive higher real wages—presumably as compensation for their misery.

Declining cities seem also to have been unhappy during the past, but in 1940, the cities that were prone to future decline earned outsized incomes, both nominal and real. The industrial cities of the Midwest may have reported lower happiness levels, but their residents were getting richer as a result. As transportation cost declines freed industry from the Great Lakes and the coal mines, we shouldn't be surprised that people left less pleasant locations. Today, the residents of cities that declined aren't receiving higher nominal wages, but they do seem to be paying lower rents. As such, the unhappiness of America's declining

cities may have been compensated with higher incomes in the past and lower housing costs today.

## Chapter 3

# The Location Decisions of Multi-establishment Firms in a Spatial and Industrial Network

### 3.1 Introduction

Multi-unit firms comprise an overwhelming percent of economic activity. Like their single-unit counterparts, multi-unit firms choose locations to gain productive advantages like productivity-enhancing spillovers, proximity to suppliers, or natural advantages, and for access to markets.<sup>1</sup> But when choosing locations, multi-unit firms face a more complex decision: they must weigh the management of spatial supply networks and the cannibalization of market access of other plants.

As a result, the geographies of multi-unit plants exhibit a series of unique spatial patterns. For example, firms with multiple establishments cluster,<sup>2</sup> but larger, more productive firms tend to cluster less, spreading their operations over larger distances. These patterns and

---

<sup>1</sup>See Rosenthal and Strange (2004a) for a review of the literature on agglomeration economics, and especially Rosenthal and Strange (2001); Ellison *et al.* (2010); Ellison and Glaeser (1997).

<sup>2</sup>See Holmes (2011).

the forces shaping them have a potentially large impact on the distribution of economic activity; the potential for market cannibalization may decrease the density gradient by inducing firms to disperse their plants, while intra-firm management costs of distance have the opposite effect, inducing firms to centralize plants around the firm's center. However, we lack a theoretical framework that allows us to model and evaluate how the decisions of multi-unit establishments create and respond to these forces in general equilibrium.

This paper proposes such a framework by adapting a new trade model of export platforms to a domestic context. Firms choose locations for multiple plants and products, trading off proximity to markets, suppliers, and local advantages for the cannibalization of sales and profits from other plants and the effects of distance on establishment management. These tradeoffs become complex in a spatial setting, where sales and spillovers are possible between any two locations.

Until now, the literatures that have dealt with entry of multi-unit firms has deals with this complexity by eliminating within-market proximity: in this (ubiquitous) city-as-island approach, costs of production and access to markets are city-level variables and are not based on proximity to other markets or other establishments within a city.<sup>3</sup> As a result, both within-city heterogeneity and the decision to expand or contract within cities is eliminated, and changes in cross-city proximity are ignored. This framework allows for a unified consideration of all forces potentially affecting these decisions.

Section two of this paper uses restricted-access US Census data to present a series of stylized facts regarding the geography of multi-unit firms: firms expand outward from their center like rings on a tree; more productive firms span larger distances and the area they inhabit is less densely populated by establishments; plants lose productivity as they're built further out from the firm's center; and new plants cannibalize markets of existing ones. While I find no convincing evidence of supply-chain effects, there are clear productivity losses from locating away from the firm's center. I don't find substantial support for location-effects related to density. While it remains unclear as to whether the number of products a

---

<sup>3</sup>See Holmes (2005) for an example. Zheng (2014) segments markets endogenously using sales data.

plant produces diminishes when a firm builds new plants nearby, the existing plant's sales decline.

These descriptive relationships between productivity and distance measures provide new accounts of several previously theorized agglomeration forces. Moreover, new evidence on cannibalization and the productivity costs of distance implies the existence of two undertheorized agglomeration and dispersion forces: intra-firm costs of distance, and market cannibalization. Because the urban literature generally divorces itself from geographic interconnections, urban models have trouble modeling the forces that capture these facts.

In the appendix presents a model of location decisions that matches these stylized facts. I extend a model of export platforms (Tintelnot 2014) in order to model these decisions in a tractable way in a domestic context, where proximity to markets as well as other industries affects competitiveness, and goods as well as factors of production are mobile. In the model, each firm produces a continuum of goods using inputs from other firms in other industries in a nested-CES production setting. As in Tintelnot (2014), this paper uses a within-firm analogy to the Eaton-Kortum framework, hypothesizing that each firm produces a continuum of goods, receiving location-specific productivity draws for each good and chooses locations for production of each good, trading off access to markets for underlying comparative advantage in production as well as the profits from other plants. The stochastic nature of within-firm production solves a previously complex problem, yielding closed-form solutions for the locations of production and the flow of goods between locations. I extend the Tintelnot framework into a domestic multi-industry setting, where production is networked both by geography and input-output relations, and location sizes are endogenously determined. I do not account here for within-vs-across firm sourcing decisions, which undoubtably do have a large impact on the location of the firms' production network. Section five concludes.

## 3.2 Data

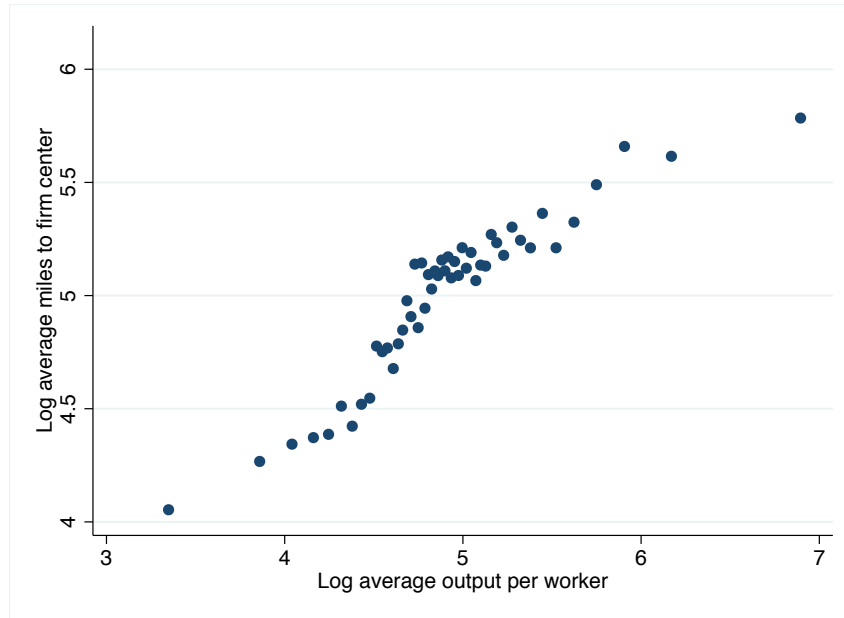
I use restricted-access establishment and product-level data on sales from all US firms responding to any bi-decadal US economic census between 1992 and 2007, and the Longitudinal Business Database, which tracks plants across years. I use address information from IRS data to match all establishments to Census tracts and to obtain latitudes and longitudes for each establishment.<sup>4</sup> For all firms, I use sales and employment at the establishment level and for manufacturing firms, I use sales at the plant level and value added at the plant level. I exclude single-employee establishments. My final sample is comprised of 5 million establishments in total, and just under 275,000 manufacturing establishments.

**Table 3.1:** *Summary statistics*

|   | Observations              | Mean     | Std. Dev. |
|---|---------------------------|----------|-----------|
|   | Mutli-unit establishments |          |           |
| <i>Sales (thousands)</i>                    | 5,007,900                 | 12,542   | 1,124,761 |
| <i>Employment</i>                           | 5,007,900                 | 41.12671 | 182.1049  |
| <i>Age</i>                                  | 5,007,900                 | 10.35    | 8.55      |
| <i>Total purchased services (thousands)</i> | 83,800                    | 3,722    | 26,122    |
| <i>Number of products</i>                   | 107,700                   | 3.98     | 5.23      |
| <i>Intermediate input intensity</i>         | 273,900                   | -0.91    | 0.71      |
|   | Distances (log miles)     |          |           |
| <i>To firm geographic center</i>            | 5,007,900                 | 4.83     | 2.08      |
| <i>To oldest establishment</i>              | 5,007,900                 | 4.28     | 2.69      |
| <i>To geographic center of new plants</i>   | 4,034,600                 | 5.04     | 2.10      |

Note. Sample includes manufacturing or business services establishments responding to Economic Censuses between 1992 and 2007 with more than one establishment, excluding sole proprietorships, for which geographic data is available or could be imputed from address records. Data on value added, intermediate input intensity, total purchased services, and number of products are taken from the subset of firms responding to the Census of Manufactures only.

<sup>4</sup>Latitude and longitude information exists in the Longitudinal Business Database starting in 2007 and Census Tract about half of establishments in each year – starting with about one third in 1992 and rising to two thirds by 2007. Using address matching across multiple observations, I am able to assign an additional tract, latitude and longitude information to an additional 40% of firms. For 20%, I use zip codes to impute tract. Only 10% of establishments cannot be traced to specific tracts using the data available. Most of these have PO Box addresses.



**Figure 3.1:** *Log average output per worker vs log average distance from firm center*

### 3.3 Firm productivity and plant dispersion

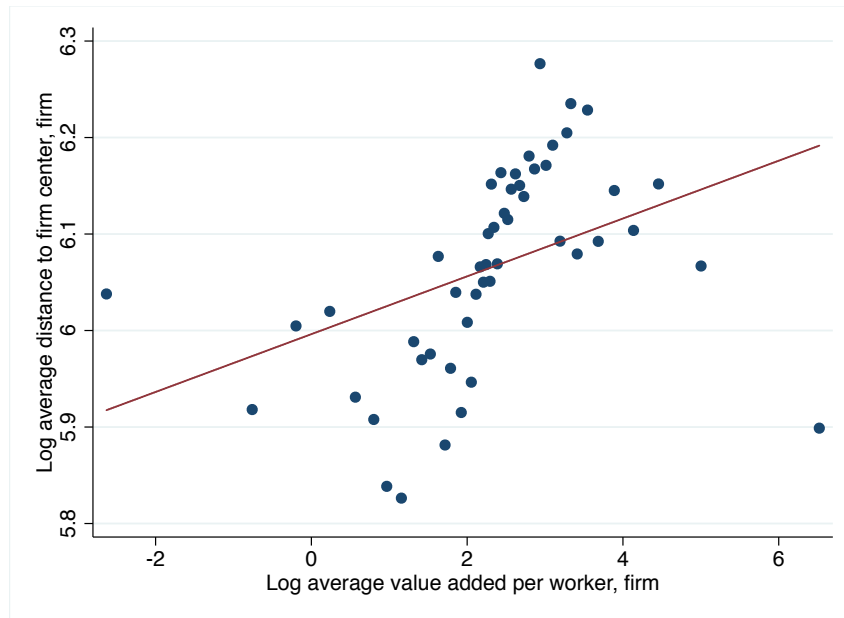
More productive firms are larger, both in terms of the number of establishments and their geographic reach. Figures 3.1-3.3 plot firm-level log output per worker, firm-level log value added per worker, and log average plant size (employees per establishment), respectively, against the log average distance from a firm's plant to its centroid. These binned scatterplots control for modal four-digit industry and year. The clear positive relationship demonstrates that more productive firms generally have more dispersed plants.

While more productive firms cover wider areas, they also have more establishments. Yet their establishments are also further away from each other on average. Figure 3.4 plots the firm-level output per worker against the establishment density, measured by establishments per square mile within the circle defined by firm's maximal radius.<sup>5</sup> The clear negative

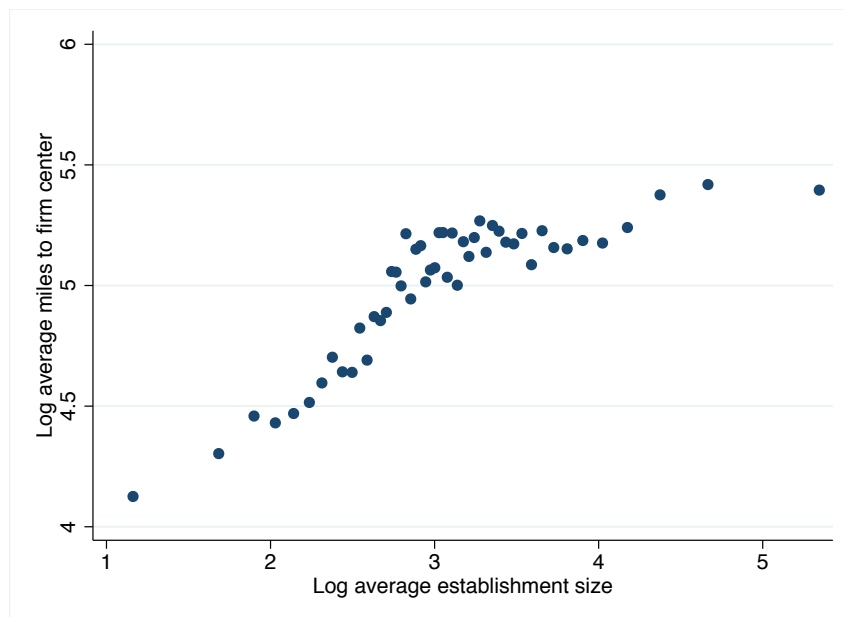
---

<sup>5</sup>Again controlling for modal four-digit industry code and year.

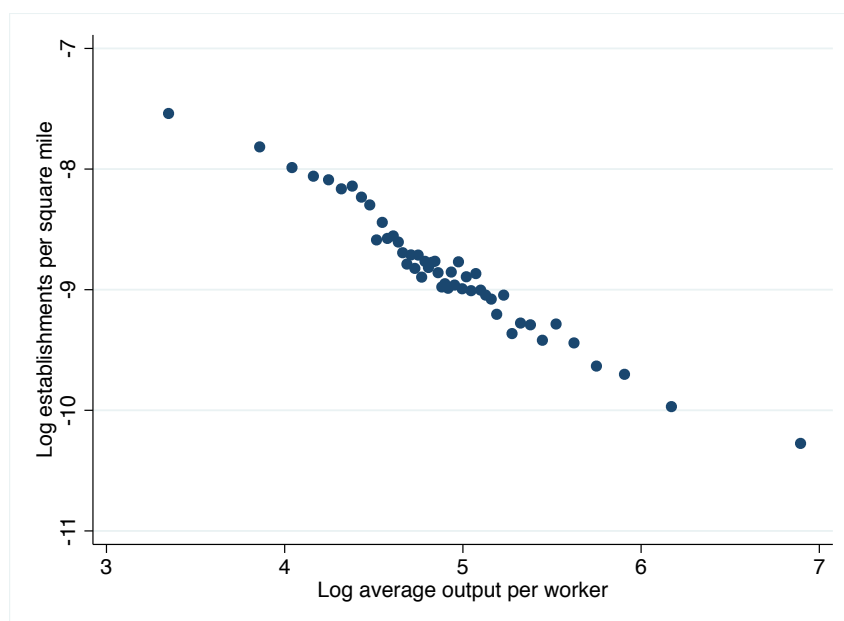




**Figure 3.2:** *Log average value added per worker vs log average distance from firm center*



**Figure 3.3:** *Log average employment per worker vs log average distance from firm center*



**Figure 3.4:** *Log average output per worker vs log average establishment density*

relationships in both figures 2a and 2b conclusively show that despite the larger number of establishments, more productive firms spread their establishments further away from each other.

From these figures alone, it remains unclear whether this correlation arises from the differential site selection of productive firms, or firm productivity is a consequence of this dispersion ((Bernard *et al.*)). However, when I consider establishment-level data I find a very different relationship between distance and productivity.

### 3.4 Evidence on internal economies of density

Firms expand radially out from their center like rings on a tree. Figure 3.4 plots the binscatter of establishment age against distance from the firm's centroid. Rather than choosing random

locations or purposefully far locations, new establishment locations very gradually and consistently push the firm's boundary and radius out.

**Table 3.2:** *Effect of distance from firm center on establishment productivity*

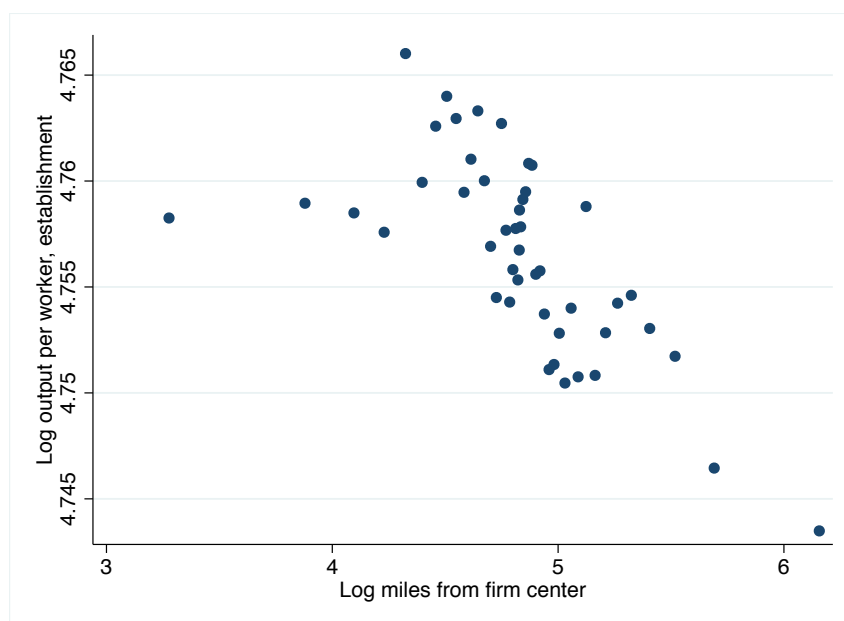
|                                    | <b>Productivity (establishment level)</b> |                        |                        |                     |                        |                        |
|------------------------------------|---|------------------------|------------------------|---------------------|------------------------|------------------------|
|                                    | (1)                                       | (2)                    | (3)                    | (4)                 | (5)                    | (6)                    |
|                                    | <b>LSPW</b>                               | <b>LVAPW</b>           | <b>LSPW</b>            | <b>LVAPW</b>        | <b>LSPW</b>            | <b>LVAPW</b>           |
| <i>Log miles from firm center</i>  | -0.0066***<br>(0.0016)                    | -0.0030***<br>(0.0025) |                        |                     | -0.0026***<br>(0.0016) | -0.0024***<br>(0.0012) |
| <i>Log miles from oldest plant</i> |   |                        | -0.0013***<br>(0.0003) | -0.0002<br>(0.0015) |                        |                        |
| <i>Plant fixed effects</i>         | No  | No                     | No                     | No                  | Yes                    | Yes                    |
| Constant                           | 9   | 5.4                    | 9.6                    | 5.1                 | 4.5                    | 4.8                    |
| Observations                       | 5,007,900                                 | 273,900                | 5,007,900              | 273,900             | 5,007,900              | 273,900                |
| R-squared                          | 0.85                                      | 0.86                   | 0.79                   | 0.82                | 0.89                   | 0.87                   |

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Note. Sample in columns one, three, and five includes all establishments that were part of multi-unit firms responding to Economic Censuses between 1992 and 2007 for which geographic data is available or could be imputed from address records Sample in columns two, four, and six includes just respondents to the Census of Manufacturers. LSPW is log output per woerks. LVAPW is log value added per worker. All regressions control for firm by industry by year fixed effects, establishment age, cubit functions of establishment latitude longitude, and full industry code fixed effects. All standard errors are clustered at the CBSA level.

Establishments that are placed further away from the firm's center are less productive. Table 3.2 reports the relationship between an establishments productivity and the establishment's distance from the firm's center<sup>6</sup> controlling for firm fixed effects and a variety of establishment controls including establishment age and cubic polynomials of latitude and longitude, industry-year fixed effects, and, importantly, firm fixed effects. No matter the measure, plants that are further away from the firm's center are less productive. Columns one and two show that doubling distance to the center reduces productivity by about 0.5%. This drops when measuring firm center as the location of the oldest establishment in columns three and four. Columns five and six add establishment fixed effects. Here

<sup>6</sup>Again I alternate between output per worker and value added per worker as productivity measures. I measure distance to the firm center in columns 1 and 2 as the distance from the geographic center of firm, averaging over all firm establishments (unwieghted), and then in columns 3 and 4 as distance to the firm's oldest establishment.



**Figure 3.5:** *Log output per worker vs log distance from firm center, firm-year and age fixed effects*

variation in distance to center comes from firm expansions. Remarkably, these estimates remain in the range of columns one through four. Figure 3.5 plots a binned scatterplot of establishment productivity against distance from firm center controlling for firm fixed effects and plant age.

In contrast to the evidence presented in Figures 3.1 through 3.3, the evidence in Figure 3.5 and Table 3.2 is consistent with a productivity-cost of distance. While firms that are observed to be more productive span larger distances, plants are less productive when they're further from the firm's center, and become less productive when the firm moves away from them. Taken together, this evidence suggests that firms bare costs of spreading out, but assess these costs differentially according ex-ante differences.

### **3.5 Market cannibalization**

Table 3.3 examines the effect of proximity to new establishments on existing establishment output. Columns one and two report the effects of new plant proximity on the number of products per establishment at existing plants, controlling for establishment fixed effects (and establishment age), without and with firm-year fixed effects as well. While adding firm-year fixed effects appears to imply that decreasing the average distance to new establishments decreases the number of products at the establishment level, the estimates are noisy, and large effects in either direction cannot be ruled out. The integer nature of the outcome variable and low establishment-level variation over time reduces the power of this test.

Columns three and four report the effects of distance to new establishments on sales at existing establishments with analogous controls. In column four, adding firm-year fixed effects makes brings the large, positive and significant relationship between distance and sales into focus: halving the distance to existing plants reduces the output of those plants by between 0.5% and 1.5%.

### **3.6 Headquarter services and intra-firm trade**

Broadly speaking, this apparent productivity cost of distance is consistent with several well-known hypothesis. Distant establishments may be difficult to monitor or increase managerial costs such as the cost of knowledge transfer (Giroud, 2011; Bahar, 2014). Distance may also increase that supply chain management costs within the firm increase with distance (Holmes, 2011; Alcacer and Delgado, 2013).

Table 3.4 searches for evidence of the latter hypothesis, assessing the effects of distance on intermediate good usage and business services usage. It's important to note that the origin of these services is unavailable in the data, so Table 3.4 conflates general changes to the supply chain with within-firm changes.

**Table 3.3: Market cannibalization**

|                                  | Number of products  |                     | Log sales             |                        |
|----------------------------------|---------------------|---------------------|-----------------------|------------------------|
|                                  | (1)                 | (2)                 | (3)                   | (4)                    |
| <i>Log miles from new plants</i> | -0.0001<br>(0.0217) | -0.0380<br>(0.0219) | 0.0154***<br>(0.0005) | -0.0050***<br>(0.0005) |
| <i>Firm-year fixed effects</i>   | No                  | Yes                 | No                    | Yes                    |
| Constant                         | 0.97                | 5.4                 | 9.6                   | 5.1                    |
| Observations                     | 107,700             | 107,700             | 4,034,600             | 4,034,600              |
| R-squared                        | 0.97                | 0.97                | 0.95                  | 0.97                   |

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Note. Sample in columns two and four includes all establishments that were part of multi-unit firms responding to Economic Censuses between 1992 and 2007 for which geographic data is available or could be imputed from address records Sample in columns one and three includes just respondents to the Census of Manufacturers. All regressions control for establishment and year fixed effects. All standard errors are clustered at the CBSA level.

Columns three, six, and nine report intermediate good usage, total services purchased, and total services purchased per employee controlling for firm-industry-year fixed effects. Contrary to the supply-chain hypothesis, intermediate input usage appears to increase with distance to the firm's center, although large negative effects cannot be excluded.

However, columns three and nine point to evidence that establishments far from the firm center may be substituting within-firm services for external services. While service usage does decline substantially, these establishments are also smaller. Moreover, larger firms tend to use higher levels of internally-provided business services.<sup>7</sup> In Column nine, while I find no evidence of negative effects on the value of services per employee at the establishment level. Doubling the distance from firm center increases external services consumption by one percent. Taken together, this establishment-level evidence neither immediately rejects supply chain productivity effects nor supports such hypotheses but is consistent with theory-of-the-firm explanations for the decrease in productivity.

<sup>7</sup>See (Antràs and Helpman, 2003)

**Table 3.4: Internal agglomeration economies**

|                                   | Log intermediate input usage |                        |                    | Log total services used |                        |                         | Log total services per employee |                        |                       |
|-----------------------------------|------------------------------|------------------------|--------------------|-------------------------|------------------------|-------------------------|---------------------------------|------------------------|-----------------------|
|                                   | (1)                          | (2)                    | (3)                | (4)                     | (5)                    | (6)                     | (7)                             | (8)                    | (9)                   |
| <i>Log miles from firm center</i> | -0.0080***<br>(0.0013)       | -0.0031***<br>(0.0001) | 0.0009<br>(0.0041) | 0.1979***<br>(0.0061)   | -0.1229***<br>(0.0051) | -0.09751***<br>(0.0139) | 0.5450***<br>(0.0036)           | -0.0443***<br>(0.0033) | 0.0111***<br>(0.0128) |
| <i>Industry by year FEs</i>       | No                           | Yes                    | Yes                | No                      | Yes                    | Yes                     | No                              | Yes                    | Yes                   |
| <i>Firm x ind. by year FEs</i>    | No                           | No                     | Yes                | No                      | No                     | Yes                     | No                              | No                     | Yes                   |
| Constant                          | -4.9                         | -2.6                   | -1.4               | -7.4                    | -0.7                   | 12.7                    | 3.9                             | 3                      | 4.7                   |
| Observations                      | 273,900                      | 273,900                | 273,900            | 83,800                  | 83,800                 | 83,800                  | 83,800                          | 83,800                 | 83,800                |
| R-squared                         | 0.01                         | 0.26                   | 0.79               | 0.12                    | 0.33                   | 0.64                    | 0.01                            | 0.23                   | 0.57                  |

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Note. Sample includes all establishments that were part of multi-unit firms responding to the Census of Manufacturers between 1992 and 2007 for which geographic data is available or could be imputed from address records. Intermediate input usage is one minus the value added over the total output. All regressions control for year fixed effects, establishment age, and cubic functions of establishment latitude longitude. All standard errors are clustered at the CBSA level.

### 3.7 Evidence on external economies of density

A traditional story for location choice is the existence of productivity spillovers associated with density (Duranton and Puga, 2004; Combes and Duranton, 2006). Table 3.5 uses the panel nature and the multi-unit nature of the data to test for evidence of local density effects. Column one reports the baseline elasticity of establishment productivity with respect to tract-level density at 5%, controlling for establishment-level industry-year effects, age, and cubic polynomials of latitude and longitude. This result is consistent with a broad range of estimates.<sup>8</sup>

When firm fixed-effects are added in column two, the estimates drop to 1.5%, and a 1% elasticity can be excluded when, in column three, establishment fixed-effects are included. While this potentially suggests selection or sorting, Maré and Graham (2009) argue that firm fixed-effects bias results such as these downwards due to the convexity of such (potential) forces. Columns four and five attempt to address these concerns where possible by instead using pairs of plant expansions. Using only pairs of new establishments within the same firm that open in differentially dense Census tracts, column 4 reports a similarly low effect of moving into denser tracts. Column five controls for CBSA fixed effects as well and finds

<sup>8</sup>See Rosenthal and Strange (2004a) for a summary, and especially Behrens *et al.* (2010); Gaubert (2014).

**Table 3.5: External agglomeration economies**

|  | Log output per worker |                       |                     |                       |                       |                    |                    |                     |
|--|-----------------------|-----------------------|---------------------|-----------------------|-----------------------|--------------------|--------------------|---------------------|
|  | (1)                   | (2)                   | (3)                 | (4)                   | (5)                   | (6)                | (7)                | (8)                 |
| <i>Log establishment density, t</i>    | 0.0409***<br>(0.0029) | 0.0174***<br>(0.0014) | 0.0017*<br>(0.0010) |                       |                       |                    | 0.0008<br>(0.0020) |                     |
| <i>Log establishment density, CBSA</i> |                       |                       |                     | 0.0549***<br>(0.0025) | 0.0235***<br>(0.0016) | 0.0017<br>(0.0011) |                    | -0.0027<br>(0.0024) |
| <i>Firm by industry fixed effects</i>  | No                    | Yes                   | Yes                 | No                    | Yes                   | Yes                | Yes                | Yes                 |
| <i>Establishment fixed effects</i>     | No                    | No                    | Yes                 | No                    | No                    | Yes                | No                 | No                  |
| Constant                               | 8.6                   | 7.9                   | 5.6                 | 7.5                   | 7.5                   | 5                  | 5.5                | -1.3                |
| Observations                           | 5,007,900             | 5,007,900             | 5,007,900           | 5,007,900             | 5,007,900             | 5,007,900          | 129,120            | 129,120             |
| R-squared                              | 0.59                  | 0.81                  | 0.79                | 0.57                  | 0.81                  | 0.79               | 0.95               | 0.96                |

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Note. Sample in columns one through six includes all establishments that were part of multi-unit firms responding to Economic Censuses between 1992 and 2007 for which geographic data is available or could be imputed from address records. Sample in columns seven and eight use the subset of these establishments which are new establishments (younger than five years) that are part of pre-existing multi-unit firms, opening in a CBSA at least 50 miles away from any other establishments of the same firm.. All regressions control for year fixed effects, establishment age, and cubic functions of establishment latitude longitude. All standard errors are clustered at the CBSA level.

no evidence of positive productivity effects of density within cities. It should be noted that these specifications do not eliminate the potential for site selection.

Columns six through nine address the assertion of Maré *et al.* (2006) that location effects are stronger at higher levels of aggregation, repeating columns one through four at the CBSA level and find similar results.

### 3.8 Conclusion

This paper documents and explains the location decisions of multi-establishment firms, and in doing so, isolates two new agglomeration and dispersion forces: intra-firm distance costs and market cannibalization, the former of which pushes firms to centralize production while the latter induces firms to spread in a less dense manner.

I use the establishment-year level data from the universe of US multi-unit establishments in order to document the geography of these firms. I find that more productive firms are spread over larger distances. Despite having more establishments, their establishments are less densely populated. While the average distance to the firm center is larger for more



productive firms, establishments within the firm are less productive when they are located further away from the firm's center. Firms tend to expand radially out from their centers.

These facts simultaneously imply significant site selection by firms, and suggest that firms face a cost of expanding geographically, which they mitigate by keeping establishments close. These intra-firm distance costs cannot yet be pinned down by any particular story. However they appear to relax for more productive firms, which cover larger distances, and as firms gain productivity and expand, they travel the upper envelope, expanding outward in increments.

In addition, when firms expand, their existent plants suffer sales losses. This cannibalization pushes firms to centralize production in fewer establishments and to spread establishments over further distances. The data suggests that more productive firms are more elastic to this effect.

Finally, I propose the Tintelnot (2014) framework in order to embed these new forces into a model of multi-unit establishment location decisions. In the appendix, I let firms choose a set of locations from which to produce and sell a range of products. Locations differ by natural advantages, the cost of labor and proximity to upstream and downstream linkages. Observed firm productivity is a function of these agglomeration forces, ex-ante heterogeneity, and intra-firm organizational costs. Together, these differences along with random product draws give certain plants comparative advantage over others in a set of goods.

In the model, production takes place both within a spatial network and within an industrially interconnected framework that explicitly models the effects of proximity to upstream linkages. This is the geographic alternative to the standard framework in the urban literature<sup>9</sup> which models input-output relations within cities only. The estimation procedure this paper will follow will yield (1) new estimates of agglomeration forces previously modeled, including natural advantage, and local labor costs, (2) new estimates of forces modeled here (for the first time) as geographic forces, including upstream and downstream

---

<sup>9</sup>Ellison *et al.* (2007)

linkages, and (3) estimates of geographic forces not previously modeled: cannibalization and intra-firm distance costs.

# References

- AGARWAL, R. P., O'REGAN, D. and WONG, P. J. (2008). Constant-sign solutions for systems of fredholm and volterra integral equations: the singular case. *Acta Applicandae Mathematicae*, **103** (3), 253–276.
- AHLFELT, G., REDDING, S., STURM, D. and WOLF, N. (2012). The economics of density: Evidence from the berlin wall. *Working Paper*.
- ALBOUY, D. (2008). Are big cities really bad places to live? improving quality-of-life estimates across cities. *Ann Arbor*, **1001**, 48109–1220.
- ALCACER, J. and DELGADO, M. (2013). Spatial organization of firms and location choices through the value chain.
- ALESINA, A., DI TELLA, R. and MACCULLOCH, R. (2004). Inequality and happiness: are europeans and americans different? *Journal of Public Economics*, **88** (9), 2009–2042.
- ALLEN, T. and ARKOLAKIS, C. (2013). *Trade and the Topography of the Spatial Economy*. Tech. rep., National Bureau of Economic Research.
- ALONSO, W. *et al.* (1964). Location and land use. toward a general theory of land rent. *Location and land use. Toward a general theory of land rent*.
- ANTRÀS, P. and HELPMAN, E. (2003). *Global sourcing*. Tech. rep., National Bureau of Economic Research.
- BAHAR, D. (2014). Heavier than air? knowledge transmission within the multinational firm.
- BALDWIN, R. E. and OKUBO, T. (2006). Heterogeneous firms, agglomeration and economic geography: spatial selection and sorting. *Journal of Economic Geography*, **6** (3), 323–346.
- BATES, D. M. and PINHEIRO, J. C. (1998). Computational methods for multilevel modelling. *Madison, WI: University of Wisconsin*.
- BAUMEISTER, R. F. (1991). *Meanings of life*. Guilford Press.
- BECKER, G. S. (1965). A theory of the allocation of time. *The economic journal*, pp. 493–517.
- , RAYO, L. and KRUEGER, A. B. (2008). Economic growth and subjective well-being: Re-assessing the easterlin paradox. comments and discussion. *Brookings Papers on Economic Activity*, pp. 88–102.

- BEHRENS, K., DURANTON, G. and ROBERT-NICOUD, F. (2012). Productive cities: sorting, selection, and agglomeration. *Working Paper*.
- , —, — *et al.* (2010). *Productive cities: Sorting, selection and agglomeration*. Centre for Economic Policy Research London.
- BENJAMIN, D. J., KIMBALL, M. S., HEFFETZ, O. and REES-JONES, A. (2012). What do you think would make you happier? what do you think you would choose? *The American economic review*, **102** (5), 2083.
- BENTHAM, J. (1879). *An introduction to the principles of morals and legislation*. Clarendon Press.
- BERNARD, A. B., MOXNES, A. and SAITO, Y. U. (). Production networks, geography and firm performance.
- CAMERON, A. C., GELBACH, J. B. and MILLER, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, **29** (2).
- CICCONE, A. and HALL, R. (1996). Productivity and the density of economic activity. *American Economic Review*, **86** (1), 54–70.
- CLARK, A. E. and OSWALD, A. J. (1994). Unhappiness and unemployment. *The Economic Journal*, pp. 648–659.
- COMBES, P.-P. and DURANTON, G. (2006). Labour pooling, labour poaching, and spatial clustering. *Regional Science and Urban Economics*, **36** (1), 1–28.
- , — and GOBILLON, L. (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, **63** (2), 723–742.
- , —, —, PUGA, D. and ROUX, S. (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica*, **80** (6), 2543–2594.
- , —, — and ROUX, S. (2010). Estimating agglomeration economies with history, geology, and worker effects. In *Agglomeration Economics*, University of Chicago Press, pp. 15–66.
- COSTINOT, A. and VOGEL, J. (2009). *Matching and inequality in the world economy*. Tech. rep., National Bureau of Economic Research.
- DALY, M. C., OSWALD, A. J., WILSON, D. and WU, S. (2011). Dark contrasts: The paradox of high rates of suicide in happy places. *Journal of Economic Behavior & Organization*, **80** (3), 435–442.
- DAVIS, D. R. and DINGEL, J. I. (2013). The comparative advantage of cities. *Columbia University, mimeograph*.
- DEATON, A. and STONE, A. A. (2014). Evaluative and hedonic wellbeing among those with and without children at home. *Proceedings of the National Academy of Sciences*, **111** (4), 1328–1333.

- DI TELLA, R., MACCULLOCH, R. J. and OSWALD, A. J. (2001). Preferences over inflation and unemployment: Evidence from surveys of happiness. *American Economic Review*, pp. 335–341.
- DURANTON, G. and PUGA, D. (2004). *Handbook of Regional and Urban Economics*, Elsevier B.V., chap. Micro-Foundations of Urban Agglomeration Economies, pp. 2063–2117.
- EASTERLIN, R. A. (1974). Does economic growth improve the human lot? some empirical evidence. *Nations and households in economic growth*, **89**, 89–125.
- (1995). Will raising the incomes of all increase the happiness of all? *Journal of Economic Behavior & Organization*, **27** (1), 35–47.
- ELLISON, G. and GLAESER, E. (1997). Geographic concentration in us manufacturing industries: a dartboard approach. *Journal of Political Economy*, **105** (5), 889–927.
- , — and KERR, W. (2010). What causes industry agglomeration? evidence from coagglomeration patterns. *American Economic Review*, **100**, 1195–1213.
- and GLAESER, E. L. (1994). *Geographic concentration in US manufacturing industries: a dartboard approach*. Tech. rep., National Bureau of economic research.
- , — and KERR, W. (2007). *What causes industry agglomeration? Evidence from coagglomeration patterns*. Tech. rep., National Bureau of Economic Research.
- FISHER, I. (2006). *Mathematical Investigations in the Theory of Value and Prices, and Appreciation and Interest*. Cosimo, Inc.
- FLORIDA, R., MELLANDER, C. and RENTFROW, P. J. (2013). The happiness of cities. *Regional Studies*, **47** (4), 613–627.
- FOR DISEASE CONTROL, C., (CDC), P., FOR DISEASE CONTROL, C., (CDC), P. *et al.* (2001). Behavioral risk factor surveillance system survey questionnaire. *Atlanta, Georgia: US Department of Health and Human Services, Centers for Disease Control and Prevention*, pp. 22–23.
- FUJITA, M., KRUGMAN, P. and VENABLES, A. J. (1999). The spatial economy.
- GAUBERT, C. (2014). *Essays on Firms and Cities*. Ph.D. thesis, Princeton University.
- GILBERT, D. (2009). *Stumbling on happiness*. Vintage Canada.
- GIROUD, X. (2011). Proximity and investment: Evidence from plant-level data. *Available at SSRN 1958200*.
- GLAESER, E. L. (2007). *The economics approach to cities*. Tech. rep., National Bureau of Economic Research.
- and GOTTLIEB, J. D. (2008). *The economics of place-making policies*. Tech. rep., National Bureau of Economic Research.
- , — and ZIV, O. (2014). *Unhappy cities*. Tech. rep., National Bureau of Economic Research.

- and KOHLHASE, J. E. (2004). Cities, regions and the decline of transport costs\*. *Papers in Regional Science*, **83** (1), 197–228.
- , KOLKO, J. and SAIZ, A. (2001). Consumer city. *Journal of economic geography*, **1** (1), 27–50.
- and REDLICK, C. (2009). Social capital and urban growth. *International Regional Science Review*.
- , RESSEGER, M. and TOBIO, K. (2009). Inequality in cities\*. *Journal of Regional Science*, **49** (4), 617–646.
- and TOBIO, K. (2007). *The rise of the sunbelt*. Tech. rep., National Bureau of Economic Research.
- GLENN, N. D. and WEAVER, C. N. (1979). A note on family situation and global happiness. *Social Forces*, pp. 960–967.
- GOLOMB, M. (1935). Zur theorie der nichtlinear integralgleichungen, integralgleichungssysteme und allgemeinen functionalgleichungen. *Math. Zeit.*, **39**, 45–75.
- GYOURKO, J. (2009). *Understanding commercial real estate: just how different from housing is it?* Tech. rep., National Bureau of Economic Research.
- HOLMES, T. J. (2005). The location of sales offices and the attraction of cities. *Journal of Political Economy*, **113** (3), 551–581.
- (2011). The diffusion of wal-mart and economies of density. *Econometrica*, **79** (1), 253–302.
- KAHNEMAN, D. and KRUEGER, A. B. (2006). Developments in the measurement of subjective well-being. *The journal of economic perspectives*, **20** (1), 3–24.
- KYRIAKOPOULOU, E. and XEPAPADEAS, A. (2013). Environmental policy, first nature advantage and the emergence of economic clusters. *Regional Science and Urban Economics*, **43** (1), 101–116.
- LEE, Y. (2008). Geographic redistribution of us manufacturing and the role of state development policy. *Journal of Urban Economics*, **64** (2), 436–450.
- LUCAS, R. E. (2013). Life satisfaction of us counties predicts population growth. *Social psychological and personality science*, p. 1948550613494227.
- and ROSSI-HANSBERG, E. (2002). On the internal structure of cities. *Econometrica*, **70** (4), 1445–1476.
- LUDWIG, J., DUNCAN, G. J., GENNETIAN, L. A., KATZ, L. F., KESSLER, R. C., KLING, J. R. and SANBONMATSU, L. (2012). Neighborhood effects on the long-term well-being of low-income adults. *Science*, **337** (6101), 1505–1510.
- LUTTMER, E. F. (2004). *Neighbors as negatives: Relative earnings and well-being*. Tech. rep., National Bureau of Economic Research.
- MANDEVILLE, B. (1806). *The Fable of the Bees; Or, Private Vices, Public Benefits....*

- MARÉ, D. C. and GRAHAM, D. J. (2009). Agglomeration elasticities in new zealand.
- , TIMMONS, J. and ECONOMIC, M. (2006). *Geographic concentration and firm productivity*. Tech. rep., Motu Economic and Public Policy Research.
- MELITZ, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, **71** (6), 1695–1725.
- MIRRLYES, J. A. (1976). The optimal structure of incentives and authority within an organization. *The Bell Journal of Economics*, pp. 105–131.
- NOCKE, V. (2006). A gap for me: Entrepreneurs and entry. *Journal of the European Economic Association*, **4** (5), 929–956.
- OSWALD, A. J. and WU, S. (2010). Objective confirmation of subjective measures of human well-being: Evidence from the usa. *Science*, **327** (5965), 576–579.
- and — (2011). Well-being across america. *Review of Economics and Statistics*, **93** (4), 1118–1134.
- PETRIN, A., WHITE, T. K. and REITER, J. P. (2011). The impact of plant-level resource reallocations and technical progress on us macroeconomic growth. *Review of Economic Dynamics*, **14** (1), 3–26.
- RAPPAPORT, J. and SACHS, J. D. (2003). The united states as a coastal nation. *Journal of Economic growth*, **8** (1), 5–46.
- ROBACK, J. (1982). Wages, rents, and the quality of life. *The Journal of Political Economy*, pp. 1257–1278.
- ROSEN, S. (1979). Wage-based indexes of urban quality of life. *Current issues in urban economics*, **3**.
- ROSENTHAL, S. and STRANGE, W. (2001). The determinants of agglomeration. *Journal of Urban Economics*, **50** (2), 191–229.
- and — (2004a). *Handbook of Regional and Urban Economics*, North Holland, chap. Evidence on the Nature and Sources of agglomeration economies, pp. 191–229.
- ROSENTHAL, S. S. and STRANGE, W. C. (2004b). Evidence on the nature and sources of agglomeration economies. *Handbook of regional and urban economics*, **4**, 2119–2171.
- ROSSI-HANSBERG, E. (2005). A spatial theory of trade. *American Economic Review*, pp. 1464–1491.
- RUGGLES, S., SOBEK, M., FITCH, C. A., HALL, P. K. and RONNANDER, C. (1997). *Integrated public use microdata series: Version 2.0*. Historical Census Projects, Department of History, University of Minnesota.
- SACKS, D. W., STEVENSON, B. and WOLFERS, J. (2010). *Subjective well-being, income, economic development and growth*. Tech. rep., National Bureau of Economic Research.

- STEVENSON, B. and WOLFERS, J. (2008). *Economic growth and subjective well-being: Reassessing the Easterlin paradox*. Tech. rep., National Bureau of Economic Research.
- STIGLER, G. J. (1950). The development of utility theory. i. *The Journal of Political Economy*, pp. 307–327.
- SWEET, J. A. and BUMPASS, L. L. (1996). The national survey of families and households—waves 1 and 2: Data description and documentation. *Center for Demography and Ecology, University of Wisconsin-Madison* (<http://www.ssc.wisc.edu/nsfh/home.htm>).
- , — and CALL, V. (1988). *The design and content of the National Survey of Families and Households*. University of Wisconsin, Center for Demography and Ecology Madison.
- TINTELNOT, F. (2014). *Global production with export platforms*. Tech. rep., WTO Staff Working Paper.
- WHITE, T. K., REITER, J. P. and PETRIN, A. (2012). *Plant-level Productivity and Imputation of Missing Data in US Census Manufacturing Data*. Tech. rep., National Bureau of Economic Research.
- ZHENG, F. (2014). *Spatial Competition and Preemptive Entry in the Discount Retail Industry*. Tech. rep., Working paper, Harvard University.



# Appendix A

## Appendix to Chapter 1

### A.1 Proof of sorting equilibrium.

The proof will proceed as follows. First, conditional on there being multiple values of  $\eta$  for different locations, I show that higher  $\eta$  locations must in equilibrium attract higher  $\psi$  firms. I then show that differences in  $\eta$  must exist by ruling out cases in which no differences exist in the space  $S$ .

#### A.1.1 Sorting conditional on non-degenerate distribution

Assume the opposite, that there exists two locations  $i$  and  $j$  where  $i$  has a higher location specific productivity, so  $\eta(i) > \eta(j)$ , but firms at  $i$  have lower productivity  $\psi_2$  than firms at  $j$ , with productivity  $\psi_1$ , so  $\psi_1 > \psi_2$ . Rents must be such that firms in neither location wish to move. Rent is a fixed cost. It is incurred by firms of different productivities in the same way, whereas, by Proposition 2, higher  $\eta$  has differential effects on firms according to their productivity. Lower rent at  $j$  must compensate the higher productivity firm at  $j$  for decreased variable profit.  $\phi(i) - \phi(j) \geq \frac{(\rho\psi_1)^{\sigma-1}}{\sigma} (\eta(i) - \eta(j))$ , but since  $\psi_1 > \psi_2$ , this implies  $\phi(i) - \phi(j) > \frac{(\rho\psi_2)^{\sigma-1}}{\sigma} (\eta(i) - \eta(j))$ . But this violates the second incentive compatibility constraint of the firm (See equation (4)) at  $j$ ,  $\phi(j) - \phi(i) \geq \frac{(\rho\psi_2)^{\sigma-1}}{\sigma} (\eta(i) - \eta(j))$ . Therefore the firm at  $j$  could not have optimally chosen  $j$  in equilibrium.

### A.1.2 Proof of impossibility of complete non-sorting equilibrium

Next, I show that an equilibrium in which no location differs by  $\eta(i)$ , i.e. that  $\eta(i)$  is constant for any  $i \in S^n$ , can be ruled out. The proof evaluates changes in the value of  $\eta(i)$  at a convex boundary of the space  $S$ . Because at the boundary  $\eta(i)$  cannot be constant, non-sorting equilibria can be ruled out.

Again I proceed by contraction, assuming for every  $i \in S^n$   $\eta(i) = \eta(\ell)$ . Then we can write the equation for  $\eta(i)$ , according to equation (10), as

$$\eta(i) = \int_{j \in S} P(i)^{-\sigma} (1 - \xi(i)) \left[ \frac{\tau(i, j)}{\rho P(j)} \right]^{1-\sigma} h(j) \psi(j)^\sigma \cdot \eta(j) dj.$$

Pulling out the constant value of  $\eta$  and simplifying

$$P(i)^\sigma = \int (1 - \xi(i)) \left[ \frac{\tau(i, j)}{\rho} \right]^{1-\sigma} P(j)^{1-\sigma} h(j) \psi(j) dj. \quad (\text{A.1})$$

Differentiating with respect to  $i$  and substituting in using equation (12)

$$\begin{aligned} \sigma P(i)^{-\sigma-1} \nabla_i P(i) \cdot \int_{j \in S} \left[ \frac{\tau(i, j)}{\rho P(j)} \right]^{1-\sigma} h(j) \psi(j) dj = \\ P(i)^{-\sigma} \cdot (1 - \sigma) \int_{j \in S} \tau(i, j)^{-\sigma} \nabla_i \tau(i, j) P(j)^{\sigma-1} \rho^{\sigma-1} h(j) \psi(j) dj = \\ \nabla_i P(i) \cdot \int_{j \in S} \sigma P(i)^{-1} \tau(i, j)^{1-\sigma} P(j)^{\sigma-1} \rho^{\sigma-1} h(j) \psi(j) dj = \\ (1 - \sigma) \int_{j \in S} \tau(i, j)^{-\sigma} \nabla_i \tau(i, j) P(j)^{\sigma-1} \rho^{\sigma-1} h(j) \psi(j) dj \quad (\text{A.2}) \end{aligned}$$

where,

$$\begin{aligned} \nabla_i P(i) = \int_{j \in S^n} \left[ \frac{\tau(i, j)}{\rho} \right]^{1-\sigma} P(j)^{1-\sigma} \psi(j)^{\sigma-1} h(j) dj \frac{\sigma}{1 - \sigma} \cdot \\ \int_{j \in S^n} \tau(i, j)^{-\sigma} \nabla_i \tau(i, j) \rho^{\sigma-1} P(j)^{1-\sigma} \psi(j)^{\sigma-1} h(j) dj. \end{aligned}$$

Equation (A.2) can be rewritten as

$$\int_{j \in S^n} \tau(i, j)^{-\sigma} \nabla_i \tau(i, j) \left( P(j)^{1-\sigma} \psi(j)^{\sigma-1} \cdot z(i) \cdot h(j) + (\sigma - 1) P(j)^{\sigma-1} h(j) \psi(j) \right) dj = 0 \quad (\text{A.3})$$

where

$$z(i) = \frac{\int_{j \in S^n} \left[ \frac{\tau(i, j)}{\rho} \right]^{1-\sigma} P(j)^{1-\sigma} \psi(j)^{\sigma-1} h(j) dj}{\int_{j \in S^n} \sigma P(i)^{-1} \tau(i, j)^{1-\sigma} P(j)^{\sigma-1} \rho^{\sigma-1} h(j) \psi(j) dj}$$

is strictly positive over  $i \in S^n$ .

Expressing this condition by grouping locations for which  $\nabla \tau(i, j)$  is positive,  $J_2 \in S^n$ , and negative,  $J_1 \in S^n$ ,

$$\begin{aligned} \int_{J_1} \tau(i, j)^{-\sigma} \nabla_i \tau(i, j) \left( P(j)^{1-\sigma} \psi(j)^{\sigma-1} \cdot z(i) \cdot h(j) + (\sigma - 1) P(j)^{\sigma-1} h(j) \psi(j) \right) dj = \\ \int_{J_2} \tau(i, j)^{-\sigma} \nabla_i \tau(i, j) \left( P(j)^{1-\sigma} \psi(j)^{\sigma-1} \cdot z(i) \cdot h(j) + (\sigma - 1) P(j)^{\sigma-1} h(j) \psi(j) \right) dj \end{aligned}$$

But in the limit as  $i$  approaches the edge of the space, the definition of  $\tau(i, j)$  reduces the set of  $J_2$  to zero. While the right hand side of the equation must go to zero, every term on the left hand side is by definition positive, making the sum itself positive. This condition is therefore a contradiction when evaluated at a convex boundary of the space.

### A.1.3 Partial sorting equilibria

At this point, it may still be possible for some group of locations  $j$  to have equivalent location productivity potential  $\bar{\eta} = \eta(j)$ . In this case, some subset of firms are made indifferent between all locations  $j$ , landowners charge  $\bar{\phi} = \phi(j)$  constant rents, and density is constant  $\bar{h} = h(j)$  across all locations  $j$ . Marginal costs of production  $w(j)$  may differ across these locations as may the price index  $P(j)$  and therefore market access  $\frac{\eta(j)P(j)}{w(j)^{1-\sigma}}$ .

Although complete non-sorting has been ruled out, this partial sorting equilibrium necessitates a many-to-one match of some set of firm types to locations. In this case, the mapping  $\psi(\eta)$  is discontinuous and the rent curve  $\phi(\eta)$  is kinked at  $\bar{\eta}$ .

## A.2 Landowner density and firm choice decisions

A firm's willingness to pay for a space  $i$  depends on its variable profit at  $i$  vs that at any other location  $j$  as well as the rents that firm would face at  $j$ . In particular, a firm with productivity parameter  $\psi$  will be willing to pay rent for space  $i$  according to the function:

$$WTP(\psi, i) = \min_j \left\{ \psi^{\sigma-1} (\eta(i) - \eta(j)) + \phi(j) \right\} \quad (\text{A.4})$$

where  $\phi(j)$  is the rent faced by firms at location  $j$  and  $\eta$  is the location-specific productivity parameter. More productive firms derive higher profits from any location, but the rent schedule derived from the IC constraints (equations (3) and (4)) makes it such that firms are unwilling to locate at less productive locations, because the rent savings are outweighed by the productivity loss, and unwilling to locate at more productive locations, because the higher rents must at least offset increased profits.

The IC constraints previously derived ensure rents are such that no firm wishes to move, but this does not immediately imply that the firm with the highest willingness to pay for a particular space is the one assigned to that space. The latter condition would ensure landowners find it optimal to provide density to the firm type matched to their location.

The following argument shows that these IC constraints do ensure just that: assuming a matching function of firms to locations and a schedule of rents supporting the firms incentives to locate, the firm with the highest willingness to pay for a given space is the firm matched to that location.

I first choose to compare the willingness to pay for space  $i$  of firm  $\psi(\eta(i))$  with any firm  $\psi < \psi(\eta(i))$ .

Choosing specifically the location  $k$  to which it is assigned, the willingness to pay of such a firm is less than or equal to the difference between its variable profit at  $i$  and  $k$  plus rent at  $k$ :

$$WTP(\psi, i) = \min_j \left\{ \psi^{\sigma-1} (\eta(i) - \eta(j)) + \phi(j) \right\} \leq \psi^{\sigma-1} (\eta(i) - \eta(k)) + \phi(k) \quad (\text{A.5})$$

However, rearranging the terms in the upper IC constraint for this firm, rent at this location, which is the willingness to pay for  $i$  of the firm  $\psi(\eta(i))$ , or  $\phi(i)$  must be greater than the right hand side of the above equation:

$$WTP(\psi, i) = \min_j \left\{ \psi^{\sigma-1} (\eta(i) - \eta(j)) + \phi(j) \right\} \leq \psi^{\sigma-1} (\eta(i) - \eta(k)) + \phi(k) \leq \phi(i).$$

An identical argument (omitted) is made for firms above  $\psi$ , using the lower IC constraint. I therefore conclude that the firm with the highest willingness to pay for location  $i$  must be the firm matched to  $i$  using the function  $\psi(\eta(i))$  in an equilibrium where rents support firm location incentives.

Alternatively, we can write the equilibrium landowner maximization problem

$$\phi(i) = \max_{\psi} \left\{ \min_j \left\{ \psi (\eta(i) - \eta(j)) + \phi(j) \right\} \right\} \quad (\text{A.6})$$

Again dropping the minimization, i.e. reducing each firm's decision to one between their assigned location and  $i$ , and substituting the equilibrium condition for rents (recall)

$$\psi(\eta) = \left[ \left( \frac{\partial \phi(\eta)}{\partial \eta} \right) \right]^{\frac{1}{\sigma-1}}$$

we can rewrite the landowners problem as

$$\phi(i) = \max_{\psi} \left\{ \psi(\eta)^{\sigma-1} (\eta(i) - \eta(j)) + \int_{\eta}^{\eta(j)} \psi(\eta)^{\sigma-1} d\eta \right\} \quad (\text{A.7})$$

Finally, using the implicit function theorem (since  $\eta(j)$  changes with  $\psi$ ) and a change of variables:

$$\phi(i) = \max_{\psi} \left\{ \psi(\eta)^{\sigma-1} (\eta(i) - \eta(\psi(j))) + \int_{\psi(\eta)}^{\psi(\eta(j))} \psi(\eta)^{\sigma-1} \left( \frac{\partial \psi}{\partial \eta} \right)^{-1} d\psi \right\}$$

we can evaluate the first order condition:

$$0 = (\sigma - 1)\psi(\eta)^{\sigma-2} (\eta(i) - \eta(\psi(j))) - \psi(\eta)^{\sigma-1} \left( \frac{\partial \psi}{\partial \eta} \right)^{-1} + \psi(\eta)^{\sigma-1} \left( \frac{\partial \psi}{\partial \eta} \right)^{-1}$$

which is solved where  $\eta(\psi(i)) = \eta(\psi(j))$  or by choosing the firm that has already been assigned to that location.

## A.3 Proof of Existence and conditions for uniqueness

### A.3.1 Existence

First, equations (10)-(12) can be written as a system of nonlinear Hammerstein equations of the second kind.

Equation (11) in particular can be rewritten as

$$a(i) = \int \tau(i, j)^{1-\sigma} F(a(j), b(j)) dj \quad (\text{A.8})$$

where the functions  $a(j) = \frac{\eta(j)}{P(j)^{-\sigma}}$  and  $b(j) = P(j)^{1-\sigma}$  are defined by an integral equation with a kernel of  $\tau(i, j)^{1-\sigma}$  and a nonlinear function

$$F(a(j), b(j)) = \rho^{\sigma-1} \cdot g(\psi(a(j) \cdot b(j)^{\frac{-\sigma}{1-\sigma}})) \psi'(a(j) \cdot b(j)^{\frac{-\sigma}{1-\sigma}}) \cdot \psi(a(j) \cdot b(j)^{\frac{-\sigma}{1-\sigma}})^{2-2\sigma} \cdot a(j) \cdot b(j)^{\frac{1}{\sigma-1}},$$

In turn, the function  $b(j)$  can be used to similarly rewrite equation (12) as

$$b(i) = \int \tau(i, j)^{1-\sigma} G(a(j), b(j)) dj \quad (\text{A.9})$$

where the function  $b(j)$  is defined over an integral equation with a kernel of  $\tau(i, j)^{1-\sigma}$  and a nonlinear function

$$G(a(j), b(j)) = b(j) \cdot \rho^{\sigma-1} \psi(a(j) \cdot b(j)^{\frac{-\sigma}{1-\sigma}})^{\sigma-1} g(\psi) \cdot \psi'(a(j) \cdot b(j)^{\frac{-\sigma}{1-\sigma}})$$

Together, equations (A.9) and (A.10) constitute a system of nonlinear Hammerstein equations of the second kind. Agarwal *et al.* (2008) show the existence for such a system on four conditions they refer to as (C1)-(C4). I will show that conditions (C1)-(C4) are satisfied in the system described by (A.9) and (A.10), and therefore that an equilibrium exists.

First, by assumption, the kernel of this system of equations  $\tau(i, j)^{1-\sigma}$  is continuous and non-negative. This satisfies condition (C1). Note that  $F(a(j), b(j))$  and  $G(a(j), b(j))$  accept only non-negative arguments and must be non-negative everywhere, and are continuous and closed for any non-negative elements  $a(j)$ ,  $b(j)$ , since, by equation (10),  $\psi(\eta(i))$  and  $\psi'(\eta(i))$  are continuous functions. This satisfies condition (C2). The distribution of firm productivities is bounded by some maximum  $\bar{\psi}$  and  $g(\bar{\psi})$ , maximum density for some firm type, is finite. Notice that because the rent at each location is finite, the derivative function  $\psi'(\eta)$  must be finite everywhere and we can define  $\bar{\psi}'$  as its maximum.

Next, define

$$d(i) = \sup \int_S \tau(i, j) dj$$

and

$$w_{F1}(a(j)) = a(j), w_{F2}(b(j)) = b(j)^{\frac{1}{\sigma-1}}, w_{G1}(a(j)) = 1, w_{G2}(b(j)) = b(j) q = \rho^{\sigma-1} \cdot \bar{\psi} \cdot \bar{\psi}' \cdot g(\bar{\psi})$$

Then, by construction,  $G(a(j), b(j))$  and  $F(a(j), b(j))$  are both less than  $q \cdot w_{F1}(a(j)) \cdot w_{F2}(b(j))$  and  $q \cdot w_{G1}(a(j)) \cdot w_{G2}(b(j))$ , respectively, which are finite. This satisfies conditions (C3).

Finally, note that

$$\alpha > dq \alpha^{\frac{2\sigma-1}{\sigma-1}}$$

for

$$\alpha < (dq) \frac{1 - \sigma}{2\sigma - 1},$$

fulfilling (C4). This completing the sufficient conditions for existence of a nontrivial solution to the system.

### A.3.2 Uniqueness

In the general case, the equilibrium of this model will not be unique with respect to the mapping of firms to locations. However, two forms of restrictions on the kernel, or the trade costs, and therefore restrictions on the geography underlying the model, admit a single equilibrium.

Golomb (1935) shows that the system of equations in (A.9) and (A.10) has a unique solution if the following Lipschitz condition is satisfied:

$$(F(a_1, b_1) - F(a_2, b_2))^2 \leq k_1^2 \left( (a_1 - a_2)^2 + (b_1 - b_2)^2 \right) \quad (\text{A.10})$$

$$(G(a_1, b_1) - G(a_2, b_2))^2 \leq k_2^2 \left( (a_1 - a_2)^2 + (b_1 - b_2)^2 \right),$$

for any possible  $a_1, b_1, a_2, b_2$ , and some  $k_1, k_2$  such that

$$k_1^2 + k_2^2 < \lambda$$

where  $\lambda$  is the smallest eigenvector from the kernel defined by

$$K(i, j) = \int_{S^n} (\tau(i, r) \tau(j, r))^{1-\sigma} dr$$

### A.3.3 Stability

The equilibrium is point-wise locally stable if no small group of entrepreneurs or workers can increase their welfare by moving to a different location and no group of landowners



can increase profits by changing the amount of density they provide. I show that any equilibrium with a one-to-one, continuous matching  $\psi(\eta)$  is stable.

First, no group  $\epsilon$  of landowners  $i \in \epsilon$  can improve profits by adjusting the density of firms at their locations by some fixed amount  $\gamma$ . Adjusting density downward lead (by Appendix 2) to a decrease in profits and to a decrease in  $\eta(i)$ . Adjusting density upwards for each landowner reduces profits by a fixed amount through increased marginal cost for each landowner, while potentially increasing  $\eta(i)$  via higher local demand. As  $\epsilon$  becomes small, the effect of local changes due to  $\epsilon\gamma$  on  $\eta(i)$  go to zero, since  $\eta(i)$  is defined with respect to all points  $j \in S^n$ . For any change  $\gamma$  in density, there is an  $\epsilon$  small enough such that  $\Delta\eta(i) < c(h(i) + \gamma) - c(h(i))$ . Since this is true for any change in density  $\gamma$ , no arbitrarily small group of landowners can increase profits by deviating in their density provision.

Next, no small group of firms and workers can improve their profits by moving to another location. Firms moving to a new location pay higher rents for the increase in density but may benefit from better market access, as higher demand for their goods from other firms and workers in their deviating group drive up variable profits at the new location. Following Allen and Arkolakis (2013), this cannot be the case when  $\frac{d\pi(i)}{dh(i)} < 0$  which is the case if  $\frac{d\eta(i)}{dh(i)} < -\frac{d\phi(i)}{dh(i)}$ . Intuitively, because  $\eta(i)$  is defined with respect to the entire space, smaller groups of firms have an increasingly smaller effect on  $\eta(i)$ , while the negative effect of density on profits via the direct impact on  $\phi(i)$  is constant.

Worker's real wages are always equalized across locations, so no independent move by workers can improve their utility.

## A.4 Conditions for a mono-centric equilibrium

Urban models and some parts of the new economic geography literature (Fujita Krugman Venables 1999) have traditionally assumed a single, central business district. No such organization of space is assumed in this model. Although the model may yield a single center of economic activity, other equilibria may have multiple “centers”. Because cities often

have a single business center and because the literature has so often assumed geographies with exogenous centers, in this appendix I examine conditions under which economic activity must necessarily be mono-centric in the space.

The space  $S$  is mono-centric if there are no troughs of economic activity. As shown in sections 2 and 3 above,  $\eta(i)$  is a sufficient statistic for economic activity at any location. So conditions that exclude local minima in  $\eta(i)$  are sufficient conditions to guarantee that any equilibria is mono-centric. For the remainder of this appendix, I refer to local maxima and minima when I discuss the curvature of the function  $\eta(i)$  in the space  $S$ .

To exclude local minima in  $\eta(i)$  is to exclude any equilibria for which, for some  $i \in S$ ,  $\nabla \eta(i) = 0$  and  $\nabla^2 \eta(i) < 0$ . To simplify the computation we rewrite

$$\eta(i) = P(i)^{-\sigma} a(i) \quad (\text{A.11})$$

where  $a(i) = \int_{j \in S} \tau(i, j)^{1-\sigma} z(j) dj$  and  $P(i) = \int_{j \in S} \tau(i, j)^{1-\sigma} y(j) dj$ .

The first order condition is therefore

$$\frac{\partial (P(i)^{-\sigma})}{\partial i} = -\frac{\partial (a(i))}{\partial i} \frac{P(i)^{-\sigma}}{a(i)}. \quad (\text{A.12})$$

Next, I evaluate the second derivative of  $\eta(i)$ . At a local minima, this must be positive.

$$\frac{\partial^2 \eta}{\partial i^2} = \frac{\partial^2 (P(i)^{-\sigma})}{\partial i^2} a(i) + 2 \frac{\partial (P(i)^{-\sigma})}{\partial i} \frac{\partial (a(i))}{\partial i} + \frac{\partial^2 (a(i))}{\partial i^2} P(i)^{-\sigma} \quad (\text{A.13})$$

With further substitution of the first order condition, equation (A4.3) can be rewritten as

$$\frac{\partial^2 \eta}{\partial i^2} = \frac{\partial^2 (P(i)^{-\sigma})}{\partial i^2} a(i) - 2 \frac{P(i)^{-\sigma}}{a(i)} \left( \frac{\partial (a(i))}{\partial i} \right)^2 + \frac{\partial^2 (a(i))}{\partial i^2} P(i)^{-\sigma} \quad (\text{A.14})$$

or

$$\begin{aligned} \frac{\partial^2 \eta}{\partial i^2} = & (1 - \sigma) \int_{j \in S} \left( (-\sigma) \tau(i, j)^{-\sigma-1} \nabla \tau(i, j) + \tau(i, j)^{-\sigma} \nabla^2 \tau(i, j) \right) y(j) dj \cdot a(i) - 2 \frac{P(i)^{-\sigma}}{a(i)} \left( \frac{\partial(a(i))}{\partial i} \right)^2 \\ & + (1 - \sigma) \int_{j \in S} \left( (-\sigma) \tau(i, j)^{-\sigma-1} \nabla \tau(i, j) + \tau(i, j)^{-\sigma} \nabla^2 \tau(i, j) \right) z(j) dj \cdot P(i)^{-\sigma} \end{aligned}$$

which is greater than zero if

$$-\sigma \tau(i, j)^{-\sigma-1} \nabla \tau(i, j) + \tau(i, j)^{-\sigma} \nabla^2 \tau(i, j) > 0$$

or

$$\nabla^2 \tau(i, j) > \sigma \frac{\nabla \tau(i, j)}{\tau(i, j)} \quad (\text{A.15})$$

When transportation costs are sufficiently convex relative to the elasticity of substitution, no local minima are possible. Intuitively, space in local minima is worse than spaces in any direction. Moving in any direction brings firms closer to their own local demand but further from centers of demand on the other side of the local minima. Convex transportation costs and substitutability of goods jointly make the such tradeoffs of proximate markets for further markets too dear. The result is that distributions with local centers of activity and local valleys cannot be supported.

It should be noted that this is a sufficient but not a necessary condition for single-peaked equilibria. If this condition is violated, the first term is negative while the second is positive. Depending on the functions  $z(j)$ ,  $y(j)$ ,  $a(j)$ , and  $P(i)$ , the second derivative may still be negative everywhere, in some places, or nowhere.

Using the flexible functional form for transportation costs  $\tau(i, j) = (1 + \|i - j\|)^d$ , where transport costs depend on the distance between points  $i$  and  $j$  and the parameter  $d$ , which reflects the extent of convexity in the transportation cost, the condition in (A.15) is

$$d > \frac{1 + \sqrt{1 + 4\sigma}}{2} \quad (\text{A.16})$$

## A.5 Incorporating productivity spillovers

The framework laid out in Section 2 can be expanded to incorporate a flexible form of productivity spillovers. In this section I present a model where locations differ by an endogenous location productivity amenity. The index strategy combined with functional form assumptions regarding the relationship between exogenous and endogenous firm productivity allow productivity spillovers to be incorporated into location-specific productivity such that the equilibrium conditions of Section 2 hold with only slight modifications.

As before, firms sell differentiated goods at a markup over marginal cost to all locations  $j \in S^n$ . However, firm productivity, now denoted by  $\tilde{\psi}$  is now endogenously defined by the firm's location. Firm variable profit at location  $i$  can be expressed as

$$r(\psi, i)/\sigma = \int_{S^n} \frac{P(i)^{-\sigma} \cdot (1 + \xi(i))^{1-\sigma} \cdot \tau(i, j)^{1-\sigma} R(j)}{P(j)^{1-\sigma} \rho^{1-\sigma} \psi(i)^{1-\sigma}} \frac{R(j)}{\sigma} dj$$

where firm productivity is a function of exogenous firm productivity  $\psi$  and location-specific productivity spillovers

$$\psi(i) = f(\psi, s(i))$$

The location-specific productive amenities  $s$  is a function of the density, productivity, and distance of other firms.

$$s(i) = f(H, \Psi, D)$$

where  $D$  is a the (exogenous) distance function between locations,  $\Psi$  is the (endogenous) mapping of firm productivities to (all) locations, and  $H$  is the (endogenous) function governing densities at all points  $j \in S$ .

A sufficient condition for isomorphism between this model and the model in Section 2 is for firm productivity and location productivity spillovers to be multiplicatively separable:

$$\psi(i) = \psi \cdot s(i).$$

Note that this model conforms to the standard agglomeration model when  $\psi = 1$  for every firm. When both  $\psi$  and  $s(i)$  are variable, more productive firms will experience larger effects from the same value of  $s(i)$ , which is an feature of other models in the literature (Gaubert 2013) and suggested by empirical evidence (Combs et al 2012).

Under the above assumption, firm variable profits at  $i$  can be expressed as

$$r(\psi, i) / \sigma = \psi \cdot \tilde{\eta}(i)$$

where  $\tilde{\eta}(i) = s(i) \cdot \eta(i)$ . Landowner, firm, and worker optimal decisions follow as before, now as a function of  $\tilde{\eta}$  rather than  $\eta$ .

Note that  $s(i)$  is determined endogenously. Because  $\eta$  is endogenous, this does not affect equations (7)-(10). However, the mapping of location productivity  $\tilde{\eta}(i)$  to locations is now altered:

$$\tilde{\eta}(i) = s(i) \int_{j \in S} P(i)^{-\sigma} (1 - \zeta(i)) \left[ \frac{\tau(i, j)}{\rho P(j)} \right]^{1-\sigma} h(j) \psi(j)^\sigma \cdot \eta(j) dj.$$

In order to solve for an equilibrium, a solution must be provided for  $\tilde{\eta}$  as well as  $s(i)$  in conjunction with the unaltered equation for the price index. Following the proof in A3.1, a non-trivial solution to a system of three integral equations of the second kind, provided a specific functional form for  $s(i)$  is determined, can be shown to exist. As with  $\eta(i)$ , restrictions on the value of spillovers will be necessary in order to ensure conditions (C1)-(C4) are met.

## Appendix B

# Appendix to Chapter 2

### B.1 Data Appendix to Glaeser, Gottlieb and Ziv, “Unhappy Cities”

#### B.1.1 BRFSS

Throughout this paper, we follow the literature in measure happiness using self-reported survey data on subjective well-being (SWB). We use a large national survey, the Behavioral Risk Factor Surveillance System (BRFSS), conducted by the Centers for Disease Control and Prevention (CDC), which asks individuals to report on their own life satisfaction using a discrete response scale.

The CDC (2005-2010) has conducted BRFSS surveys annually since 1985, in order to study risk factors for various diseases. This is a large, nationally-representative survey, involving more than 350,000 respondents in over two thousand counties annually.

The BRFSS survey is administered by individual states via telephone interviews. The interviews are collected via computer-assisted phone calls to randomly selected landlines.<sup>1</sup> During our sample period of 2005 to 2009, the survey covers all 50 states and Washington, DC.<sup>2</sup> Individuals report their county to the interviewer, and we drop observations where

---

<sup>1</sup>CDC provides weights to adjust for differences in phone line density across areas, but we do not use these weights.

<sup>2</sup>Puerto Rico, Guam and the U.S. Virgin Islands are also included, but we drop the three territories.

county is not reported.

Based on the self-reported county, respondents live in 367 metropolitan statistical areas (MSAs) and non-metropolitan regions.<sup>3</sup> When we examine temporal patterns in the data, we restrict the sample to the 177 MSAs with at least 200 respondents in each year.

The life satisfaction question we use has been a part of the BRFSS “core” since 2005. Core questions are asked in every interview with minor exceptions. In 2009, the life satisfaction question was not asked in less than 5% of BRFSS surveys, which is approximately the same percent unasked of similar questions in the survey. This number is slightly lower in other years. Responses to LSATISFY of “refused” and “unsure” are treated as missing responses and dropped from the dataset.

One might be concerned that individual SWB is reported on a discrete scale, with values whose interpretation is not obvious. When we summarize one area’s happiness as a linear average of these discrete values, the resulting summary is undoubtedly a noisy and imperfect measure area-level happiness. We cannot solve this problem, but Stevenson and Wolfers (2008) find that more sophisticated methods yield results that are extremely highly correlated across countries (correlations are regularly above 0.99) with results from this method.

We standardize each year’s data separately, with respect to the overall mean and standard deviation for the survey year in question.

One wave of the BRFSS may actually be administered in two different years (e.g., the 2009 wave interview respondents from January 2009 through January 2010). The year fixed effects  $\gamma_t$  that we estimate represent the survey wave as opposed to the actual year of the interview.

The concern about systematic differences in individual SWB is not merely hypothetical.

---

<sup>3</sup>We use the county FIPS code to assign the respondent to a metropolitan area. We use the Office of Management and Budget’s definitions of metropolitan areas from 1999 (which correspond to data from the 2000 Census). We use Primary Metropolitan Statistical Areas (PMSAs) rather than Consolidated Metropolitan Statistical Areas (CMSAs), where applicable. We classify respondents in New England, according to their New England Consolidated Metropolitan Statistical Area (NECMA) rather than PMSA or CMSA. We classify all respondents not living in an MSA, PMSA, or NECMA as part of one “non-metropolitan region” for their state (e.g., “non-metropolitan Texas”).

On the contrary, a large body of research has documented regular patterns based on age, sex, income, life events, and other demographic characteristics.<sup>4</sup> To the extent that people sort across areas based on these same characteristics, our estimates of area-level happiness will be biased.

A small percentage of survey respondents refuse to respond to one or more of the demographic questions asked. The total fraction refusing to answer, unsure of, or not being asked at least one demographic question of interest is about 2.3

The controls for children's characteristics deserve further elaboration. While the survey nearly always has information about the number of children in the household, more detailed information is available for only one randomly selected child. In most states during most years, the BRFSS asks about the age of one randomly selected child in the household, as well as the respondent's relationship to that child.<sup>5</sup> We therefore create indicator variables for four age ranges of the randomly selected child, and six categories for the respondent's relationship.<sup>6</sup> The omitted group for these questions is respondents with no children. We add a separate dummy variable indicating respondents with children in state-years when no question was asked about a child's age.

In the working paper version, appendix Table 2 reports the coefficients on the controls in this regression, when run on our full sample of 1,574,361 respondents across five waves of BRFSS. For the most part, these coefficients are consistent with findings in the previous literature, and robust to the inclusion or exclusion of area fixed effects. In column 1, we include only the basic demographic controls discussed above. We find that age has an important influence on subjective well-being, as estimated by a fifth-order polynomial in age. On average, men are 0.036 standard deviations less happy than women. There are strongly

---

<sup>4</sup>e.g. Sacks, Stevenson, and Wolfers (2010)

<sup>5</sup>The survey is divided into core questions and modules, the latter of which each state individually elects whether to ask in their phone interviews. Individual states sometimes add additional questions on their own. None of the questions we focus on are module or state questions in any year, except for the age of one randomly selected child.

<sup>6</sup>In the 2006 survey, the age of the child is not recorded, but is imputed from the reported birthdate. In 2007, the age is recorded in the BRFSS in months, and we round this down to an integer number of years.



significant differences across races, with whites reporting the highest average well-being.

The most significant correlates of happiness in column 1 are education level and marital status. Education has one of the largest impacts on individual responses, with a range of nearly half a standard deviation from high school dropouts to college graduates. But bear in mind that this regression does not control for individual-level income, which may mediate this relationship somewhat. Marital status is also extremely important, with married individuals half a standard deviation happier than single or divorced respondents. Those reporting being separated are one-sixth of a standard deviation less happy than singles or divorcees.

Our estimates of the relationship between happiness and the presence of children in the household differ from previous findings. The existing literature has generally found a significant negative association between happiness and having children, especially young children.<sup>7</sup> In the BRFSS data, however, there seems to be a more complex relationship. This regression allows us to compute the connection between a respondent's subjective well-being and the presence of children with various characteristics in the household. To calculate the complete relationship, we need to add the coefficients for the appropriate number of children (one, two, three or more), the age of the randomly selected child (one of four categories, or unknown), and the respondent's relationship to the randomly selected child. For all of these characteristics, the coefficients presented in Appendix Table 2 are expressed relative to the omitted group of respondents with no children in the household.

Parents in a one-child household are, on average, anywhere from 0.01 standard deviations less happy than similar respondents with no child to 0.07 standard deviations happier, depending on the child's age. Older children appear to be associated with less happiness, all else equal, with 11-17-year-olds having a coefficient 0.076 standard deviations below

---

<sup>7</sup>The negative relationship between children—especially young children—and parents' happiness is widely accepted in the literature. Di Tella, MacCulloch, and Oswald (2001) report increasingly negative coefficients on life satisfaction in the EuroBarometer as the number of children increases (table A1). This finding dates back at least to Glenn and Weaver (1979), who find the negative coefficient to be largest for children under 5 years old in the General Social Survey (Table 1). The closest finding to ours is Clark and Oswald (1994), who estimate a negative effect of having one child relative to no children, and insignificant negative effects of having two or more children compared with none (Tables 2 and 3). They do not report results controlling for children's ages.

0-1-year-olds. We find increasingly positive coefficients as the number of children increases, with a bump of 0.04 standard deviations for a second child and a further 0.01 standard deviation gain with a third child or beyond.

These benign or positive relationships between children and happiness disappear if the respondent is the child's guardian but not the biological parent. Grandparents, foster parents, and unspecified other relatives have very strong negative coefficients, which wipe out the (otherwise positive) associations with most categories of number and age of children. In other specifications (not reported), we interact the number or age of children with the respondent's marital status or relationship with the random child. These regressions tend to confirm that the positive correlation between children and respondents' well-being is concentrated among married couples and respondents who are the child's biological parents, while the other groups tend to have negative associations between the presence of children and their own well being.

Even without these interactions, our data suggest a more complex relationship than that previously found between subjective well-being and the presence and age of children. These correlations are sensitive to the relationship between the children present and the individual in question. Nevertheless, it is unlikely that the inclusion of controls for relationship with the child fully explains the difference between our results and the negative coefficients on children's presence reported in other papers. The cases of non-parental relationship status are probably not sufficiently prevalent to explain the aggregate negative associations found in other datasets.

Subsequent regressions in Appendix Table 2 add controls for the respondent's economic situation. In column 2, we add dummies for labor force status. With employed individuals as the omitted group, we find that self-employment is associated with a 0.036 standard deviations more well-being, while the unemployed are 0.44 to 0.57 standard deviations less happy than employed workers. Retirees are 0.02 standard deviations less happy than workers, controlling for age, and those unable to work are 0.7 standard deviations less happy than workers. Including labor force status controls has only a modest impact on the

coefficients on other demographics, with the notable exception of the indicator for being black. This dummy reverses signs, from -0.025 in column 1 to 0.01 in column 2.

Column 3 adds controls for reported income categories, in addition to of the previous characteristics. These dummies show that happiness increases monotonically in income, with a range of 0.6 from the omitted category (less than 10,000 *per year*) to the highest – income category (above 75,000 per year). Because income is correlated with many of the other covariates, its inclusion dramatically shifts some of the coefficients on other variables, including education, unemployment, race and marital status, relative to their levels in column 2.

### **B.1.2 Aggregate Data**

Our aggregate data about the metropolitan and non-metropolitan areas in the country come from various sources. These data mostly come from the National Historical Geographic Information System (Minnesota Population Center, 2004), which compiles data from the U.S. Census. We obtain these data at the county level and consolidate them using the same metropolitan area definitions from 1999 as we use for the BRFSS. We obtain a number of quality of life measurements from Albouy (2008), and geographic data from Rappaport and Sachs (2003).

## **B.2 Data on Movers from the National Survey of Families and Households**

The NSFH is a probability sample survey of 13,017 respondents in 9,643 households plus an oversampling of minority and single-family households and households with step-children. It is a longitudinal study with three waves, the first between 1987 and 1988, the second between 1992 and 1994, and the third wave between 2001 and 2002 (Sweet and Bumpass 1996; Sweet, Bumpass, and Call 1988; Trull and Famularo 1996).

We use data from the first two waves of the NSFH. In both waves, the data contains in-

formation on family and personal characteristics of individuals and on individual subjective well-being. In particular, the NSFH asks: “First taking things all together, how would you say things are these days?” Respondents may choose to respond on a 1 to 7 scale, 1 being very unhappy and 7 being very happy. The summary statistics from this survey are shown in Panels B and C of Appendix Table 1 in the working paper version. For our regressions, we normalize the responses to have mean zero and unit variance.

To examine the relationship between changes in subjective well-being and changes in geographic location, we need to match the longitudinal NSFH data to geographic data. Because the geographic locations of survey respondents are considered confidential, we can’t link individual responses to the names of the counties or PMSAs in which those individuals reside. However, the NSFH provided us with a match between survey respondent case IDs and certain geographic characteristics (“geomerge”).<sup>8</sup> For each wave, for each publically available observation, the NSFH provided a corresponding dataset with the observation case ID number and the characteristics of the respondent’s county and PMSA. While we can’t link individual respondents to named geographic locations, we can link individuals with the relevant characteristics of their counties and PMSAs in each wave. Included in our match are census data on county and PMSA population, education, and income, other geographic amenities like crime statistics and temperature, and the county and PMSA fixed effects on subjective well-being that we estimated previously using the BRFSS.

With the geographic characteristics from both NSFH waves, we are able to isolate the population of NSFH respondents who moved counties or PMSAs. In NSFH2, 2,395 respondents report moving cities since NSFH1. Using our matched dataset, we find 1,939 respondents who both answered the question on subjective well-being and have different county characteristics for NSFH1 and NSFH2, denoting a change in the respondent’s county of residence. Of that group, we similarly find 1,480 respondents to have moved to a new

---

<sup>8</sup>We are extremely grateful to Larry Bumpass, Jack Solock, Charles Fiss, and the Center for Demography of Health and Aging University of Wisconsin-Madison for generously conducting this geomerge for us and providing us with the data. The use of these geographically merged, but not individually identified data was approved by the Institutional Review Board at the National Bureau of Economic Research.

PMSA.

Our analysis focuses on the relationship between the changes in reported subjective well-being of this population and the changes in the respondents' county and PMSA characteristics. We run regressions of the form

$$\Delta y_i = \tau + \psi \Delta \hat{u}_i + \phi \Delta X_i + \epsilon_i \quad (\text{B.1})$$

across individuals who move. The coefficient  $\psi$  identifies the relationship between changes in area-level happiness and changes in individual happiness, possibly controlling for changes in other covariates (at the area or individual level) between the two observations, captured in  $\Delta X_i$ .

## Appendix C

### Appendix to Chapter 3

In this section, I build a model which accounts for the potential forces behind these facts. Proximity to markets and suppliers, agglomeration forces, and natural advantages all affect firm location decisions and explains how firms choose how many plants to build and where to build them, and the flow of goods across industries and space.

The model is a spatial model. Unlike typical urban models, shipping costs of intermediate and final goods within and across cities are modeled explicitly, and therefore are allowed to define the firms' location decision. In addition, this allows the model to capture firms' decisions about how to space their production establishments, and the model predicts when and to where firms expand and from where they relocate.

The novel approach of this paper is an adaptation of Tintelnot (2014) to a domestic production setting, which likewise solves a complex spatial location decision of firm export platforms. Here, as in Tintelnot (2014), each firm produces a continuous of goods. For each good, firms draw location-specific productivities and decide on a location or locations of production. There fixed costs to establishing plants as well as iceberg transportation costs for trading inputs and outputs across locations.

## C.1 Environment

### C.1.1 Agents

The model consists of workers and managers, the latter of whom are organized into firms. Workers are freely mobile across locations, receive a location-specific wage and consume at their locations. Managers' locations and production assignments are determined by their firms, who optimize firm-wide profits but distribute them as shares to managers in proportion to their production. Managers consume locally.<sup>1</sup>

### C.1.2 Geography and timing

Firms and workers pay fixed cost to locate at one or more locations  $i$  from a fixed set of  $N$  locations. Between any two locations  $n_1$  and  $n_2$ , all goods pay a transport cost  $1 < \tau(n_1, n_2) < \infty$ . Following Tintelnot (2014), firms will choose locations, then receive location-product specific productivity draws and determine the extent of production at each location.

### C.1.3 Demand for final goods

There are a set of  $K - 1$  industries. Within each industry, a set of firms  $\Omega_{K-1}$  produce a continuous of goods (or varieties) indexed by  $v \in (0, 1)$ . All these varieties are nested within a single industry. Final demand at each location  $i$  for goods in industry  $k$  is driven by the partial utility function, common to workers and managers:

$$U_k = \beta_{f,k} \cdot \left( \int_{\Omega_1} \int_0^1 q_i(\omega, v)^{(\sigma-1)/\sigma} dv d\omega \right)^{\sigma/(\sigma-1)}$$

where  $\sigma > 1$  is the elasticity of substitution and  $\beta_{f,k} \in (0, 1)$  is a demand weight for industry  $k$ .

---

<sup>1</sup>This series of assumptions is not necessary but guarantees balanced trade and simplifies the algebra significantly.

In addition to demand from consumers, firms face demand for their goods from producers who use their products as intermediates. Demand for intermediates will be derived in the following sub-section.

#### C.1.4 Production

A firm in an industry  $k$  at location  $i \in Z$  assembles a specific product  $v$  using an industry-specific Cobb-Douglas production function

$$F_k(L, I_k) = \psi(\omega, v, k, i, Z) \cdot L^{\alpha_k} \cdot I_k^{1-\alpha_k} \quad (\text{C.1})$$

where  $\alpha_k$  is an industry-specific Cobb Douglas parameter;  $\psi(\omega, v, k, i, Z)$  is a firm-location-product specific efficiency defined as

$$\psi(\omega, v, k, i, Z) = \frac{\lambda(i, k) \xi(\omega) \epsilon(v, i)}{\gamma(Z, i)}$$

that can be further broken down into a firm-specific efficiency  $\xi(\omega)$ , industry- location-specific efficiency  $\lambda(i, k)$ ,  $\gamma(i, Z)$  the firm-level efficiency costs of location set  $Z$ , and a product-location specific efficiency  $\epsilon(v, i)$ , the latter of which is drawn from a Frechet distribution; and the industry-specific bundle of intermediates  $I_k$ , is defined as

$$I_k = \prod_{k' \in K} I_{k',k}^{\beta_{k',k}} \quad (\text{C.2})$$

where  $I_{k',k}$  is a bundle of intermediates from industry  $k'$  used by industry  $k$ , composed of a CES bundle of varieties from all firms within industry  $k'$ ,

$$I_{k',k} = \left( \int_{\Omega_{k'}} \int_0^1 q_i(\omega, v)^{(\sigma-1)/\sigma} dv d\omega \right)^{\sigma/(\sigma-1)}$$

and  $\beta_{k',k}$  is the Cobb-Douglas parameter of inputs from industry  $k'$  for industry  $k$  such that  $\sum_{k'} \beta_{k',k} = 1$ .



Some attention should be given to the determinants of productivity,  $\psi$ . Firm-specific efficiency  $\zeta(\omega)$  governs ex-ante heterogeneity among firms. Industry-location specific efficiency  $s(i, k)$  is to be interpreted as a natural advantage of a location. In addition, if  $s$  is endogenously set as function of the number of other firms in industry  $k$ , it should be interpreted as a productivity spillover. product-location specific efficiency  $\epsilon$  will generate the Eaton-Kortum-style ricardian competition within the firm. Finally,  $\gamma(i, Z)$  is the efficiency cost of location  $i$  conditional on the firm's chosen set  $Z$ . Because there is no headquarters in this model, I will assume this cost is a function of the distance between  $i$  and the geographic center of set  $Z$ .

The price of one bundle of intermediates in industry  $k$  at location  $i$  is

$$P_{int}(i, k) = \prod_{k' \in K} \left( \frac{P_{int,k'}(i, k)}{\beta_{k',k}} \right)^{\beta_{k',k}} \quad (C.3)$$

where  $P_{int,k'}(i, k)$ , the price of the corresponding partial bundle at location  $i$  is

$$P_{int,k'}(i, k) = \left( \int_{\Omega_{k'}} \int_0^1 p(\omega, v)^{1-\sigma} dv d\omega \right)^{\frac{1}{1-\sigma}}$$

and where  $p(\omega, v, i)$  is the price at location  $i$  of variety  $v$  from firm  $\omega$ .

Because of transport costs between locations, the cost of such a bundle will vary across locations. The final cost of a good will vary across locations according to the cost of intermediates, the cost of transport of final goods to market, and location-specific, firm-specific, and product-specific productivities.

Because for each product  $v$  each plant draws a random productivity draw, the costs to serving market  $j$  from market  $i$  are distributed as

$$\begin{aligned} & Pr \left( \frac{\left( \frac{P_{int}(i, k)}{\alpha_k} \right)^{\alpha_k} \cdot \left( \frac{w(i)}{(1 - \alpha_k)} \right)^{1 - \alpha_k}}{\gamma(i, Z)^{-1} \cdot \lambda(i, k) \cdot \xi \cdot \tau_k(i, j)^{-1}} < c \right) \\ &= 1 - \exp \left( - \left( \frac{\left( \frac{P_{int}(i, k)}{\alpha_k} \right)^{\alpha_k} \cdot \left( \frac{w(i)}{(1 - \alpha_k)} \right)^{1 - \alpha_k}}{\gamma(i, Z)^{-1} \cdot \lambda(i, k) \cdot \xi \cdot \tau_k(i, j)^{-1}} \right)^{-\theta} c^\theta \right) \end{aligned} \quad (C.4)$$

where  $\tau_k(i, j)$  is the industry-specific iceberg transport costs incurred when selling from point  $i$  to point  $j$ . As in the Eaton-Kortum framework, because I have assumed a Frechet distribution, the price at location  $j$  of a variety of the firm  $\omega$ 's product will be

$$\begin{aligned} & p(\omega, j) = \\ & \kappa^{\frac{1}{1-\sigma}} \xi(\omega)^{-\theta} \left( \sum_{l \in Z} \left( \frac{\gamma(l, Z)}{\lambda(l, k)} \right)^\theta \cdot \left( \left( \frac{P_{int}(i, k)}{\alpha_k} \right)^{\alpha_k} \cdot \left( \frac{w(i)}{(1 - \alpha_k)} \right)^{1 - \alpha_k} \right)^{-\theta} \cdot \tau_k(l, j)^{-\theta} \right)^{-\frac{1}{\theta}} \end{aligned} \quad (C.5)$$

where  $\kappa = \Gamma \left( \frac{\theta + 1 - \sigma}{\theta} \right) \left( \frac{\sigma}{\sigma - 1} \right)^{1 - \sigma}$  and  $Z \in N$  are the set of locations chosen by the firm for production. The firm's sales to market  $j$  will be

$$s(\omega, j) = p(\omega, j)^{1 - \sigma} \cdot \frac{Y(j, k)}{P(j, k)^{1 - \sigma}} \quad (C.6)$$

where  $Y(j, k)$  is the total demand at location  $j$  for goods from industry  $k$ , final and intermediate. Finally, we can represent the total revenue at location  $i$  is simply the above equation summed over all locations  $j$

$$r(\omega, i, Z) = \frac{\kappa^{\frac{1}{1-\sigma}}}{\xi(\omega)^{-\theta}} \cdot \sum_j \frac{Y(j, k)}{P(j, k)^{1-\sigma}} \cdot \frac{\left( \frac{\left( \frac{P_{int}(i, k)}{\alpha_k} \right)^{\alpha_k} \left( \frac{w(i)}{(1-\alpha_k)} \right)^{1-\alpha_k}}{\lambda(i, k) \cdot \tau_k(i, j)^{-1} \cdot \gamma(i, Z)^{-1}} \right)^{-\theta}}{\left( \sum_{l \in Z} \left( \frac{\left( \frac{w(l)}{(1-\alpha_k)} \right)^{1-\alpha_k} \left( \frac{P_{int}(l, k)}{\alpha_k} \right)^{\alpha_k}}{\lambda(l, k) \gamma(l, Z)^{-1} \cdot \tau_k(l, j)^{-1}} \right)^{-\theta} \right)^{\frac{\theta+1-\sigma}{\theta}}} \quad (C.7)$$

Equation (7) demonstrates how the agglomeration and dispersion forces in the model operate. In addition to the determinants of firm productivity  $\psi$ , marginal costs are lower where the price of labor is lower and where intermediates are lower. The former is a dispersion force while the latter, proximity to upstream linkages, pushes firms to coagglomerate across industries. Transportation costs to markets increase marginal costs, acting as an agglomeration force. Finally, market cannibalization occurs through the denominator of equation (7): additional plants reduce the revenues of a given plant by reducing the set of markets to which the existing plant has a comparative advantage in servicing among all plants within the firm.

## C.2 Worker location decision

Workers provide one unit of labor inelastically to firms at their location. They demand one unit of housing inelastically at their location, and consume a basket of final goods. Housing is supplied by a competitive fringe of developers at an increasing marginal cost:

$$c(0) = 0, \quad c'(0) = 0, \quad c'(L(l)) > 0, \quad c''(L(l)) > 0$$

The total mass of workers choosing a location will be determined by the demand for labor there.

### C.3 Firm location decisions

Firms must choose the set  $Z$ , the number of plants and their location. Locating closer to their suppliers or in high productivity areas lowers production costs. Locating closer to markets lowers market-to-market costs of their products. The latter force pushes firms to disperse while the former force pushes firms to centralize production in low-cost areas. Both forces push towards co-agglomeration of linked industries. Firms must pay a fixed cost to open establishments. This also pushes production towards centralization. Finally, new establishments reduce the market access of existing establishments, as they cannibalize their markets. This force concentrates production in fewer locations and pushes establishments further away from each other. This cannibalization effect has been previously hypothesized and documented in the literature (Holmes 2011), but for which, prior to the Tintelnot (2014) framework, no general equilibrium model has been established.

Firms choose a set of locations  $Z$  which maximize their expected profits, which is variable profits – a proportion of expected sales at each location, minus the fixed costs  $f(i)$  associated with that number of locations:

$$Z(\xi, f) = \operatorname{argmax} \left\{ \sum_{i \in Z} \frac{1}{\sigma} E(s(i, \omega, Z)) - \sum_{i \in Z} f \cdot w(i) \right\}.$$

### C.4 Equilibrium

An equilibrium is defined by an  $N \times K$  set of price equations  $P(i, k)$  and demands  $Y(i, k)$  for each industry-location pair, a set of location-specific wages  $w(i)$ , that satisfy (1) a spatial equilibrium for workers, (2) labor market clearing across all markets, (3) a set of prices and location sets  $Z(\omega)$  that satisfy the firm's profit maximization condition, and (4) a balanced trade condition.

#### Worker spatial equilibrium

Because workers are freely mobile, they must be indifferent between locations, and wages at all locations are determined by this indifference condition.

$$\frac{w(l) - c(L(l))}{P(l)} = w_r = 1 \quad (\text{C.8})$$

where  $w_r$ , the real wage, is a constant normalized to 1.

#### C.4.1 The spatial-industrial network

Summing the above expression across all firms all feeder industries  $k'$  gives price index for the intermediates in industry  $k$

$$P_{int}(j, k) = \prod_{k' \in K} \beta_{k', k}^{-\beta_{k, k'}} \cdot \kappa^{\frac{\beta_{k, k'}}{1-\sigma}} \cdot \left( \int_{\Omega_{k'}} \xi(\omega)^{\sigma-1} \left[ \sum_{l \in Z} \left( \frac{\gamma(l, Z)}{\lambda(l, k) \tau_k(l, j)} \right)^\theta \left( \frac{P_{int}(i, k)}{\alpha_k} \right)^{\alpha_k - \theta} \left( \frac{w(i)}{(1 - \alpha_k)} \right)^{1 - \alpha_k - \theta} \right]^{-\frac{\sigma-1}{\theta}} dg(\xi) \right)^{\frac{\beta_{k, k'}}{1-\sigma}} \quad (\text{C.9})$$

Following Tintelnot (2014), I define  $\rho_Z^\xi$  as the share of firms with productivity  $\xi$  that choose a location set  $Z$ . The trade flows between locations  $i$  and  $j$  as

$$X(i, j) = \int_{\xi} \sum_{Z' \in Z} \rho_{Z'}^\xi E(s(i, j, \xi, Z')) dg(\xi) \quad (\text{C.10})$$

The total amount produced at location  $i$   $Y(i) = \sum_j X(i, j)$  must equal the total labor bill plus the fixed costs and the total firm profits from  $i$

$$Y(i) = \sum_j X(i, j) = w(i) \cdot L(i) + \frac{1}{\sigma} \cdot \int_{\xi} \sum_{Z' \in Z^i} \rho_{Z'}^\xi E(s(i, j, \xi, Z')) dg(\xi) + f \cdot w(i) \quad (\text{C.11})$$

### C.4.2 Labor market clearing

Finally, for the labor market to clear, the labor used in each location must sum to the total amount of labor in the economy

$$\bar{L} = \sum_{j \in N} \left( L(j) + f \cdot \int_{\tilde{\zeta}} \sum_{Z' \in \tilde{Z}} \rho_Z^{\tilde{\zeta}} dg(\tilde{\zeta}) \right) \quad (\text{C.12})$$

Together, the labor market clearing condition in (12), the goods market clearing condition, and the  $N \times K$  price equations represented in equation (10), combined with the satisfaction of the spatial equilibrium condition for workers in (9) and the firm optimal location and price conditions constitute an equilibrium.