Urban Producer Theory

by Tracy Charles Waldon

B.S. in Economics, June 1985, Montana State University
M.S. in Economics, December 1987, University of Wisconsin - Madison

A dissertation submitted to

The Faculty of
The Columbian College of Arts and Sciences
of The George Washington University
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

August 31, 2013

Dissertation directed by

Anthony M. Yezer
Professor of Economics

The Columbian College of Arts and Sciences of The George Washington University certifies that Tracy Charles Waldon has passed the Final Examination for the degree of Doctor of Philosophy as of June 6, 2013. This is the final and approved form of the dissertation.

Urban Producer Theory

Tracy Charles Waldon

Dissertation Research Committee:

Anthony Yezer, Professor of Economics; Dissertation Director

Paul Carrillo, Assistant Professor of Economics; Committee Member

Donald Parsons, Professor of Economics; Committee Member

Dedication

I dedicate this dissertation to my wife, Julie Corwin, for the unwavering support she provided during my journey.

Acknowledgments

I wish to thank the CoStar Group for generously providing data on property rents in the Cleveland CBSA and the GIS Research Laboratory at the University of Southern California for providing access to the WebGIS application.

Abstract of Dissertation

Urban Producer Theory

Urban producer theory introduces a production function which incorporates congestion in production with inputs possessing a quality component that influences productivity. These features yield cost-minimizing behavior in which firms respond to higher space rent by increasing the quality of the inputs used in production. This behavior generates demand-side sorting of high quality inputs into high rent areas. The prediction of sorting based on input quality is tested on attorneys employed in the Cleveland CBSA. Evidence of the sorting into high rent areas of attorneys based upon the national ranking of the law school attended is found. A 1% increase in rent leads to a 1.26% to 2.89% increase in the number of the highest quality attorneys employed in high rent districts. Ability sorting poses a significant risk in biasing the measurement of agglomeration economies based on wages.

Table of Contents

Dedication **iii**

Acknowledgments **iv**

Abstract of Dissertation **v**

List of Figures **viii**

List of Tables **xi**

1 Introduction **1**

2 Urban Producer Theory **3**
2.1 The Urban Producer Model with Separable Input Quality 5
2.2 The Urban Producer Model with Embodied Input Quality 18
2.3 Summary 25

3 Testing Urban Producer Theory **27**
3.1 The Data 29
3.2 Estimation Strategy 38
3.3 Estimation Based on Distance from the Central Business District . . 42
3.4 Estimation Based on Matched Locations 44
3.5 Estimation Based on Predicted Rents 47
3.6 Summary 68

4 Implications **72**
4.1 The Intra-Urban Wage Gradient 72
4.2 The Measurement of Agglomeration Economies 79
4.3 Summary 84

References **86**

Appendices **92**
A Proofs 92

B Sensitivity Analysis of Error Structure in the Rent Estimation 107
C Sensitivity Analysis of Bandwidth in the Rent Estimation 110

Figures **111**

Tables **152**

List of Figures

2.1 The effect of ρ on the input quality gradient 111

2.2 The effect of α on the input quality gradient 112

2.3 The effect of γ on the input quality gradient 113

3.1 Density of firm size and employer size in the Cleveland CBSA 114

3.2 Location of private sector attorneys in the Cleveland CBSA 115

3.3 Location of CoStar properties in the Cleveland CBSA 116

3.4 Density of Cleveland asking rents in CoStar space data 117

3.5 Probabilities and elasticities from the distance model 118

3.6 Probabilities and elasticities using matched properties: model 1 119

3.7 Probabilities and elasticities using matched properties: model 2 120

3.8 Probabilities and elasticities using matched properties: model 3 121

3.9 Probabilities and elasticities using matched properties: model 4 122

3.10 Probabilities and elasticities using matched properties: model 4 123

3.11 Probabilities and elasticities using matched properties 124

3.12 predicted rent in the Cleveland CBSA 125

3.13 predicted rent within five miles of the Cleveland CBD 126

3.14 Probabilities and elasticities using predicted rent: model 1 127

3.15 Probabilities and elasticities using predicted rent: model 2 128

3.16 Probabilities and elasticities using predicted rent: model 3 129

3.17 Probabilities and elasticities using predicted rent: model 4 130

3.18 Probabilities and elasticities using predicted rent: model 4 131

3.19 Probabilities and elasticities using predicted rent 132

3.20 Probabilities and elasticities using predicted rent: effect of experience . . 133

3.21 Elasticities using predicted rent: effect of school location 134

4.1 Estimated wage gradients in the Cleveland CBSA 135

B.1 Probabilities and elasticities using predicted rent and alternative error structures: model 1 136

B.2 Probabilities and elasticities using predicted rent and alternative error structures: model 2 137

B.3 Probabilities and elasticities using predicted rent and alternative error structures: model 2 138

B.4 Probabilities and elasticities using predicted rent and alternative error structures: model 3 139

B.5 Probabilities and elasticities using predicted rent and alternative error structures: model 3 140

B.6 Probabilities and elasticities using predicted rent and alternative error structures: model 3 141

B.7 Probabilities and elasticities using predicted rent and alternative error structures: model 4 142

B.8 Probabilities and elasticities using predicted rent: weighted and unweighted results for model 2 143

B.9 Probabilities and elasticities using predicted rent: weighted and unweighted results for model 4 144

C.1 Probabilities and elasticities using predicted rent and alternative bandwidths: model 1 145

C.2 Probabilities and elasticities using predicted rent and alternative bandwidths: model 2 146

C.3 Probabilities and elasticities using predicted rent and alternative bandwidths: model 2 147

C.4 Probabilities and elasticities using predicted rent and alternative bandwidths: model 3 148

C.5 Probabilities and elasticities using predicted rent and alternative bandwidths: model 3 149

C.6 Probabilities and elasticities using predicted rent and alternative bandwidths: model 3 150

C.7 Probabilities and elasticities using predicted rent and alternative bandwidths: model 4 151

List of Tables

3.1 Private sector employers with more than 50 attorneys in the Cleveland CBSA 152

3.2 Highest ranked U.S. law schools 153

3.3 Most common law schools attended by Cleveland private sector attorneys 154

3.4 Spatial distribution of private sector attorneys in the Cleveland CBSA . 155

3.5 Descriptive statistics of private sector attorneys in the Cleveland CBSA 156

3.6 Information provided in CoStar property database 157

3.7 Information provided in CoStar space database 158

3.8 Descriptive statistics of CoStar property data in the Cleveland CBSA . . 159

3.9 Descriptive statistics of CoStar space data in the Cleveland CBSA . . . 160

3.10 Binary logit estimate of the difference between properties with space for lease and those without space for lease 161

3.11 Least squares estimate of the rent gradient in the Cleveland CBSA . . . 162

3.12 Ordered logit estimate of the attorney quality gradient from the distance model 163

3.13 Binary logit estimate of the difference between matched and non-matched properties 164

3.14 Ordered logit estimates of the attorney quality gradient using matched properties 165

3.15 Explanatory variables of the locally weighted regressions 166

3.16 Summary statistics from the locally weighted regressions 167

3.17 Hypothesis tests from the locally weighted regressions 168

3.18 Ordered logit estimates of the attorney quality gradient using predicted rent 169

3.19 Ordered logit estimates of the attorney quality gradient using predicted rent: effect of experience 170

3.20 Ordered logit estimates of the attorney quality gradient using predicted rent: effect of school location 171

4.1 Naive wage gradients in the Cleveland CBSA 172

4.2 Estimated wage gradients in the Cleveland CBSA 173

B.1 Ordered logit estimates of the attorney quality gradient under alternative error structures: model 1 174

B.2 Ordered logit estimates of the attorney quality gradient under alternative error structures: model 2 175

B.3 Ordered logit estimates of the attorney quality gradient under alternative error structures: model 3 176

B.4 Ordered logit estimates of the attorney quality gradient under alternative error structures: model 4 177

C.1 Ordered logit estimates of the attorney quality gradient under alternative bandwidths: model 1 178

C.2 Ordered logit estimates of the attorney quality gradient under alternative bandwidths: model 2 179

C.3 Ordered logit estimates of the attorney quality gradient under alternative bandwidths: model 3 180

C.4 Ordered logit estimates of the attorney quality gradient under alternative bandwidths: model 4 181

Chapter 1

Introduction

The relation among land rent, rent of physical space, and building height has been known since Adam Smith. The consequences of space costs for output/land ratios is also well understood. However, the consequences of space costs for input quality have been ignored. I explore the relationship between inputs to production that vary in quality and the costs of the space needed to deploy those inputs. I account for the cost of space in deploying physical inputs by realizing that most physical inputs require a minimum amount of space in order to be functional; office machines and machine tools require space for placement and operation, vehicles require parking spaces, and office workers require office space. In addition, none of these inputs is homogeneous; office machines vary in speed as well as features, vehicles vary in carrying capacity and efficiency, and office workers vary in their productivity. These relatively simple modifications to standard production theory generate interesting implications for the urban landscape. Because the cost of space, commonly referred to as rent, varies within urban areas, cost minimizing firms will vary their mix of inputs in response to differences in factor prices. The result of this substitution is that firms located where the cost of space is dear will substitute inputs that do not require space for the physical

inputs that do. In the context of labor inputs, this equates to substituting high skill workers for low skill workers, while in the context of physical capital this would equate to substituting feature-laden, high-capacity, and low-footprint machinery for simple, slow, and bulky equipment. In the face of the well-understood rent gradient that exists in most urban areas, this substitution yields a sorting mechanism whereby high quality inputs are deployed in high rent areas, commonly the central business district, and low quality inputs are deployed in low rent areas, commonly the fringe of urban areas. This generates what I refer to as the input quality gradient.

This dissertation proceeds in two parts. In chapter 2, urban producer theory is developed. Two alternative means of expressing the concept of input quality are used and in both instances cost-minimizing behavior on the part of urban producers generates an input quality gradient, among other implications. Chapter 3 proceeds to test for the major implication of urban producer theory. Do high quality inputs sort into high rent areas? The test is performed by examining whether attorneys that graduated from high quality law schools are more likely to be employed in high rent areas and less likely to be employed in low rent areas than are attorneys that graduated from lower quality law schools. The dissertation concludes by examining the implications of urban producer theory for existing lines of research in urban economics. Because much empirical research on the spatial elements of productivity relies upon wages to measure said productivity, particular care must be taken to control for the sorting of inputs based on unobserved input quality.

Chapter 2

Urban Producer Theory

This chapter develops the model of firm behavior that drives urban producer theory. The basic proposition of urban producer theory is that many inputs to production require physical space. The inputs may require a minimum amount of space and the productivity of the input may be related to the amount of physical space allocated to the input. One obvious candidate for this situation is labor. Workers require a minimum amount of space and as the amount of space allocated to them falls, their productivity is likely to fall as well. However this characteristic holds for other inputs as well. Consider the case of copiers. A copier with basic performance characteristics requires a given amount of space. However, that same space can accommodate other copiers whose performance characteristics substantially exceed those of the basic model.

Minimum space requirements can be thought of as a form of congestion. When inputs such as labor and machinery are not allocated sufficient space to function properly, the space becomes congested and productivity falls unless more space is added to the production process. Congestion is not a novel concept in economics. Färe and Svensson (1980) provide a rigorous examination of the various forms that congestion

can take in a production function. However, measurement of the congestion effect predates this treatment. The most widespread application is in transportation research and the analysis of the impact of highway infrastructure and vehicle counts on the flow of traffic. Examples of this application include Boardman and Lave (1977), Inman (1978), and Färe et al. (1982). Analysis of congestion on highways takes place in an environment where the congestion effect is an externality to the economic agents and is therefore commonly observed. However, congestion effects that are internalized by the firm have also been examined. These analyses tend to focus on the productive efficiencies of firms and the departures from economically efficient input mixes. Examples of this line of inquiry include Byrnes et al. (1984), Brockett et al. (2004), and Flegg and Allen (2009).

Urban producer theory expands on these previous analyses of congestion by introducing an additional component into production. In addition to their space requirements, inputs, even when they perform the same function in the production process, do not have the same inherent level of productivity. Materials vary in this inherent level of productivity, which I refer to as quality. In the labor context we refer to this feature as human capital. In the physical capital context this quality element will take on terms such as pages per minute for printers, FLOPS (floating point operations per second) for computers, and seasonal energy efficiency ratios for air conditioners. Even raw materials will vary in quality such as the BTUs per ton of coal.

These two characteristics of inputs create a combined factor and space price for employing the input. While the price of the input may not vary by location, the cost of deploying the input will vary as the floor area rent in a location varies. This leads to variation in the cost of employing inputs. Depending on the space requirements of an input and how that relates to its quality, it may be a cost minimizing behavior to employ inputs of a certain quality level in one location but inputs of a different

quality level in another due to the differences in the deployment costs created by different floor area rents in the two locations. Urban producer theory formalizes this intuition by introducing minimum space requirements into the production function along with a measure of input quality.

Urban producer theory applies to any producer for whom input quality is not proportional to space requirements. However, testing of the theory requires that industries be identified that meet both this requirement and which also are found located though-out a metropolitan area so that firms with similar production functions can be observed paying very different space rents. This generally implies that the products of the firm have high transportation costs so that firms are forced to locate in both high and low rent areas.

I focus on two particular modeling conventions for incorporating input quality into the production function. I refer to these forms as the separable form and the embodied form. In the separable form, input quality is treated as a separate input from the physical input in the urban production function. This leads to a situation where the firm chooses an aggregate amount of the physical input and an aggregate level of quality of those inputs. The embodied form of the urban production function incorporates input quality by assuming different classes of physical inputs each of which embodies a different level of quality as evidenced by its productivity.

2.1 The Urban Producer Model with Separable Input Quality

The separable form of the urban production function envisions a situation where firms jointly choose a quantity of the physical input and the total amount of quality of that input. Frameworks such as this are fairly common in the literature on growth. A

prime example of this would be Romer (1990) and its offspring. Within the field of urban economics, Abel et al. (2012) provides a recent example of the use of the separable form of input quality in the context of the amount of human capital in an urban area. As I will show, the separable form has an added advantage that it provides relatively tractable and concrete measures of the average quality of inputs chosen by the firm.

2.1.1 The Urban Production Function

Assume a production function that uses three inputs: N, H, and S to produce output (Q). Input S represents structure or space used in production. Input N exhibits the main characteristics of urban producer theory. It has a quality component, measured by H, that acts to enhance the productivity of N employed in production. The total quality of the physical units of N employed is given by H and the average quality of input N is given by $h = \frac{H}{N}$. A minimum amount of structure must be allocated to each physical unit of N in order for output to be produced. I use the following production function where A is a Hicks-neutral productivity factor:

$$Q = \begin{cases} A\left[\alpha N^{\rho} + \beta H^{\rho} + \phi\left(S - \gamma N\right)^{\rho}\right]^{\frac{1}{\rho}} & \text{if } S - \gamma N \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

with

$$\{\alpha, \beta, \phi, \gamma\} > 0$$

$$\alpha + \beta + \phi = 1$$

$$-\infty \leq \rho < 1$$

In the terminology of Färe and Svensson (1980), this production function is

output-prohibitive congested for factor mixes where $S \leq \gamma N$. This production function is similar to that developed by Färe and Jansson (1975) and applied by Färe et al. (1982) to analyze the relationship between highway speeds and the volume of cars on a highway. In their application, two inputs are used to produce the output. The inputs are the number of vehicles per hour, which corresponds to the physical input in my model, and highway lanes, which corresponds to structure in my model. I introduce a third input, the quality of the physical inputs, which does not have a minimum space requirement. The output analyzed by Färe et al. (1982) is average vehicle speed. This contrasts with my model where output is an aggregate value rather than an average value.

This production function possesses a number of desirable properties. It is homogeneous of degree 1 and consequently homothetic as well. This guarantees exhaustion of the marginal product. It is also concave in the inputs.[1] It should be noted that while the general form of the urban production function closely resembles the commonly used constant elasticity of substitution (CES) production function, as will be shown, the elasticity of substitution is not constant once the minimum space requirement of the physical input is introduced. Like the CES production function, the urban production function is not defined for $\rho = 0$. However, in the limit as $\rho \to 0$ the function converges to a Cobb-Douglas form, $N^{\alpha} H^{\beta} (S - \gamma N)^{\phi}$.

2.1.2 Cost Minimization

In order to understand how the minimum space requirements of the urban production function influence firm behavior, I consider the cost-minimizing behavior of the firm. An alternative would be to consider the profit-maximizing behavior of the firm. However, this approach would require additional structure to the model, such as specifying

[1] Proof of this result is found in Appendix A.

demand for the firms' output as well as entry conditions. Because the focus of this research is on the input mix chosen by the firm and the factor shares are independent of the firm's output, I only consider the cost minimizing solutions. The firm solves the following problem and generates a cost function and factor demand conditional on the level of output.

$$\min_{S,N,H} wN + gH + rS$$
$$\text{s.t. } Q = A[\alpha N^{\rho} + \beta H^{\rho} + \phi\,(S - \gamma N)^{\rho}]^{\frac{1}{\rho}}$$

The solution to the cost minimization problem generates the conditional factor demands for inputs N, H, and S. They are:

$$N(g,r,w,Q) = \frac{Q}{A}\left[\frac{\alpha}{w+\gamma r}\right]^{\frac{1}{1-\rho}} \times \left[\alpha\left(\frac{\alpha}{w+\gamma r}\right)^{\frac{\rho}{1-\rho}} + \beta\left(\frac{\beta}{g}\right)^{\frac{\rho}{1-\rho}} + \phi\left(\frac{\phi}{r}\right)^{\frac{\rho}{1-\rho}}\right]^{-\frac{1}{\rho}} \quad (2.1)$$

$$H(g,r,w,Q) = \frac{Q}{A}\left(\frac{\beta}{g}\right)^{\frac{1}{1-\rho}} \times \left[\alpha\left(\frac{\alpha}{w+\gamma r}\right)^{\frac{\rho}{1-\rho}} + \beta\left(\frac{\beta}{g}\right)^{\frac{\rho}{1-\rho}} + \phi\left(\frac{\phi}{r}\right)^{\frac{\rho}{1-\rho}}\right]^{-\frac{1}{\rho}} \quad (2.2)$$

$$S(g,r,w,Q) = \frac{Q}{A}\left[\left(\frac{\phi}{r}\right)^{\frac{1}{1-\rho}} + \gamma\left(\frac{\alpha}{w+\gamma r}\right)^{\frac{1}{1-\rho}}\right] \times \left[\alpha\left(\frac{\alpha}{w+\gamma r}\right)^{\frac{\rho}{1-\rho}} + \beta\left(\frac{\beta}{g}\right)^{\frac{\rho}{1-\rho}} + \phi\left(\frac{\phi}{r}\right)^{\frac{\rho}{1-\rho}}\right]^{-\frac{1}{\rho}} \quad (2.3)$$

The firm's cost function is:

$$C(g,r,w,Q) = \frac{Q}{A}\left[g\left(\frac{\beta}{g}\right)^{\frac{1}{1-\rho}} + r\left(\frac{\phi}{r}\right)^{\frac{1}{1-\rho}} + (w+r\gamma)\left(\frac{\alpha}{w+\gamma r}\right)^{\frac{1}{1-\rho}}\right] \times \left[\alpha\left(\frac{\alpha}{w+\gamma r}\right)^{\frac{\rho}{1-\rho}} + \beta\left(\frac{\beta}{g}\right)^{\frac{\rho}{1-\rho}} + \phi\left(\frac{\phi}{r}\right)^{\frac{\rho}{1-\rho}}\right]^{-\frac{1}{\rho}} \quad (2.4)$$

The form of the conditional factor demands and cost function provide a indication of the underlying mechanics that distinguish the urban production function from standard production functions. The only difference between the functions above and those generated by CES production is the term γr which appears additively with the price of the physical input. The term $w + \gamma r$ represents the cost of deploying a physical input with the minimum amount of space it will require to be productive. All of the results of urban producer theory are driven by this wedge between the factor price of the physical input and the deployment cost of the physical input.

Following Blackorby and Russell (1989), the Morishima elasticity of substitution can be defined as:

$$M_{ij}(Q,p) = \frac{p_i C_{ij}(Q,p)}{C_j(Q,p)} - \frac{p_i C_{ii}(Q,p)}{C_i(Q,p)} \quad (2.5)$$

where $C(Q,p)$ is the cost function at output Q and factor prices p, p_i is the factor price of input i, $C_i(Q,p) = \frac{\partial C(Q,p)}{\partial p_i}$ and by Shephard's lemma is also conditional factor demand for input i, and $C_{ij}(Q,p) = \frac{\partial^2 C(Q,p)}{\partial p_i \partial p_j}$. The Morishima elasticity of substitution

of i for j measures the change in the cost minimizing ratio of inputs i and j when the factor price ratio changes due to a change in the price of input i.

Blackorby and Russell (1981) set forth the necessary and sufficient conditions under which the Morishima elasticities are constant and symmetric. The separable urban production function meets these conditions only when $\gamma = 0$, otherwise the elasticities are neither constant nor symmetric. When $\gamma = 0$, the separable urban production function takes on the standard CES form and has an elastacity of substitution of $\frac{1}{1-\rho}$. The elasticities allow me to illustrate a few additional properties of the separable urban production function.

$$M_{NH} = \left(\frac{1}{1-\rho}\right)\frac{w}{(w+\gamma r)} \tag{2.6}$$

$$M_{HN} = \frac{1}{1-\rho} \tag{2.7}$$

$$M_{SN} = \left(\frac{1}{1-\rho}\right)\frac{w}{(w+\gamma r)}\left[\frac{\left(\frac{\phi}{r}\right)^{\frac{1}{1-\rho}}}{\left(\frac{\phi}{r}\right)^{\frac{1}{1-\rho}} + \gamma\left(\frac{\alpha}{w+\gamma r}\right)^{\frac{1}{1-\rho}}}\right] \tag{2.8}$$

$$M_{NS} = \left(\frac{1}{1-\rho}\right)\frac{w}{(w+\gamma r)}\left[\frac{\left(\frac{\phi}{r}\right)^{\frac{1}{1-\rho}}}{\left(\frac{\phi}{r}\right)^{\frac{1}{1-\rho}} + \gamma\left(\frac{\alpha}{w+\gamma r}\right)^{\frac{1}{1-\rho}}}\right] \tag{2.9}$$

$$M_{SH} = \left(\frac{1}{1-\rho}\right) r \frac{\left[\left(\frac{\phi}{r}\right)^{\frac{1}{1-\rho}} r^{-1} + \gamma\left(\frac{\alpha}{w+\gamma r}\right)^{\frac{1}{1-\rho}}(w+\gamma r)^{-1}\gamma\right]}{\left[\left(\frac{\phi}{r}\right)^{\frac{1}{1-\rho}} + \gamma\left(\frac{\alpha}{w+\gamma r}\right)^{\frac{1}{1-\rho}}\right]} \tag{2.10}$$

$$M_{HS} = \frac{1}{1-\rho} \tag{2.11}$$

As expected, only when $\gamma = 0$ are the elasticities constant and symmetric. The result is an elasticity of substitution of $\frac{1}{1-\rho}$ which is the standard value for a CES production function. For the non-CES cases, several items are of note. Most noticeable

may be that two of the elasticities are constant regardless of the value of γ. These are the elasticities of substitution of H for N and H for S when the change in the factor price ratio is due to a change in the price of the quality input, H. This result occurs because changes in the price of H have proportional effects on the demand for S and N. With the exception of these two elasticities, the term $\frac{1}{1-\rho}$ has little relationship to the elasticity of substitution beyond being a constant multiplier across all of the elasticities. For example, when $\rho = 0$ in a CES production function, this indicates that the elasticities of substitution are equal to 1. That will not be the case with the separable urban production function. It should also be noted that M_{NH} and M_{HN} are increasing in ρ. This result provides a similar result to that of CES production. As the value of ρ increases the ability to substitute between the physical input and input quality increases. The impact of ρ is less transparent on the other elasticities. The production function has a comparable interpretation to the standard CES interpretations at the boundaries of the parameter space of ρ. As $\rho \to -\infty$ the production takes the Leontief form of $Q = \min\left\{\frac{N}{\alpha}, \frac{H}{\beta}, \frac{S-\gamma N}{\phi}\right\}$. As $\rho \to 1$ the production takes the linear form of $Q = \alpha N + \beta H + \phi\left(S - \gamma N\right)$.

2.1.3 The Input Quality Gradient

The average quality level of N, $\frac{H}{N} \equiv h\left(w, r, g, Q\right)$, can be expressed as a conditional factor demand as well

$$h(w,r,g,Q) \equiv \frac{H(w,r,g,Q)}{N(w,r,g,Q)}$$

$$= \frac{\frac{Q}{A}\left(\frac{\beta}{g}\right)^{\frac{1}{1-\rho}}\left[\alpha\left(\frac{\alpha}{w+\gamma r}\right)^{\frac{\rho}{1-\rho}}+\beta\left(\frac{\beta}{g}\right)^{\frac{\rho}{1-\rho}}+\phi\left(\frac{\phi}{r}\right)^{\frac{\rho}{1-\rho}}\right]^{-\frac{1}{\rho}}}{\frac{Q}{A}\left[\frac{\alpha}{w+\gamma r}\right]^{\frac{1}{1-\rho}}\left[\alpha\left(\frac{\alpha}{w+\gamma r}\right)^{\frac{\rho}{1-\rho}}+\beta\left(\frac{\beta}{g}\right)^{\frac{\rho}{1-\rho}}+\phi\left(\frac{\phi}{r}\right)^{\frac{\rho}{1-\rho}}\right]^{-\frac{1}{\rho}}}$$

$$h(w,r,g,Q) = \left[\frac{\beta(w+r\gamma)}{\alpha g}\right]^{\frac{1}{1-\rho}} \tag{2.12}$$

The average quality level is increasing in both w and r. This result is quite intuitive as these are the costs elements of deploying the physical input. As these costs rise the firm naturally shifts towards using other inputs. I define the input quality gradient as the effect of the rental rate of structures on the average input quality. It is given by:

$$\frac{\partial h}{\partial r} = \frac{\gamma}{1-\rho}\left(\frac{\beta}{\alpha g}\right)^{\frac{1}{1-\rho}}(w+\gamma r)^{\frac{\rho}{1-\rho}} \ > 0 \ \forall \ \rho \in (-\infty, 1) \tag{2.13}$$

The average quality of input N is increasing in the rental rate for structures. Note also that this effect is driven by γ. Without a minimum floor space requirement, there is no input quality gradient. This relationship is affected by a number of factors including the relative importance of the various inputs (α, β, ϕ) and the degree of substitution between the inputs (ρ).

While the input quality gradient is always positive, it may be increasing or decreasing in the rental rate for structures depending on degree of substitution between

inputs.

$$\frac{\partial^2 h}{\partial r^2} = \frac{\rho\gamma^2}{(1-\rho)^2}\left[\frac{\beta}{\alpha g}\right]^{\frac{1}{1-\rho}}\left[w+\gamma r\right]^{\frac{2\rho-1}{1-\rho}} \tag{2.14}$$

When ρ is positive, the input quality gradient is increasing in the rental rate of structures and ρ is negative, the input quality gradient is decreasing in the rental rate of structures. In other words, input quality is concave in the rental rate of structures when ρ is negative and convex when ρ is positive. This can be seen in figure 2.1.

Comparative Statics of Input Quality

I consider two effects of the parameters of the urban production function: 1) the effect on input quality, and 2) the effect on the input quality gradient. The input quality gradient determines the differences in input quality associated with changes in the rental prices of structure inputs. The steeper this slope then greater will be the differences in input quality due to differences in the rental rate of structures. As such, theses comparative statics can be used as guidance for situations in which one would be most likely to observe the deviations from the standard production function.

The role played by the input share of the physical input, α, on input quality is examined first. It is clear from equation 2.12 that input quality is decreasing in α. The effect on the input quality gradient is also negative.

$$\frac{\partial^2 h}{\partial r \partial \alpha} = -\frac{\gamma}{\alpha\left(1-\rho\right)^2}\left(\frac{\beta}{\alpha g}\right)^{\frac{1}{1-\rho}}\left(w+\gamma r\right)^{\frac{\rho}{1-\rho}} \; < 0 \; \forall \; \rho \in (-\infty, 1) \tag{2.15}$$

When the productivity of the physical input N in the production process, as measured by α, rises, the input quality gradient will be less pronounced. Recall that the input quality gradient occurs because firms substitute away from the physical input and towards quality. However, when the physical input is relatively more productive than

the quality element, the return to substituting quality for the physical input are also reduced because a larger amount of quality must be deployed in order to compensate for a fixed amount of the physical input. Therefore the cost-minimizing firm producing a constant level of output will substitute to a lesser extent and therefore the average level of quality will not rise as quickly in response to changes in the rental rate of structures.

The minimum floor space requirement, γ, clearly has a positive effect on the cost-minimizing average input quality. In addition, it has a positive impact on the input quality gradient.[2]

$$\frac{\partial^2 h}{\partial r \partial \gamma} = \frac{1}{1-\rho}\left(\frac{\beta}{\alpha g}\right)^{\frac{1}{1-\rho}} (w+\gamma r)^{\frac{\rho}{1-\rho}} \times \left[1+\left(\frac{\rho}{1-\rho}\right)\left(\frac{\gamma r}{w+\gamma r}\right)\right] > 0 \;\forall\; \rho \in (-\infty, 1) \quad (2.16)$$

The input quality gradient is steeper when the minimum floor space requirement is higher and it is easier to substitute between inputs. However, as the degree of substitution between inputs decreases (as ρ becomes more negative), the desire to substitute input quality, H, for the physical input, N, to counter the effect of the minimum floor space requirements is limited by the productivity losses due to the limited ability to substitute.

The effect of ρ on input quality depends on the parameter values.

$$\frac{\partial h}{\partial \rho} = (1-\rho)^{-2}\left[\frac{\beta\,(w+\gamma r)}{\alpha g}\right]^{\frac{1}{1-\rho}} \ln\left(\frac{\beta\,(w+\gamma r)}{\alpha g}\right) \quad (2.17)$$

The sign of this result is ambiguous. It depends on the term in the logarithm, which, while always positive, may be greater than or less than 1. When the term is greater

[2] A proof can be found in appendix A

than 1, increases in ρ will increase the average quality of inputs, h. When the term is less than 1, increases in ρ will lead to a lower average quality level. This term will be greater than 1 when $\frac{w+\gamma r}{g} > \frac{\alpha}{\beta}$. So when the 'productivity adjusted' cost of the physical input relative to the cost of quality is greater than the ratio of the importance of the physical input relative to quality then average quality is increasing in the ability to substitute between inputs. As similar, though less intuitive result holds for the effect of ρ on the input quality gradient.

These two results are illustrated in figure 2.1 which shows that for a set of fixed values of α, β, ϕ, γ, w, and g, the input quality level and its gradient can be high or low relative to other levels and gradients depending on the rental rate on structures (and its resulting influence on the deployment costs of the physical input).

Elasticity of Input Quality

It may be most instructive in gauging the significance of the input quality gradient to examine the elasticity of input quality with respect to the rental rate on structures.

$$\epsilon_{hr} \equiv \frac{\partial \ln h}{\partial \ln r} = \frac{\gamma r}{(1-\rho)(w+\gamma r)} \tag{2.18}$$

This expression has a particularly significant implication. While input quality, as demonstrated in equation (2.13), is dependent on all input prices and most parameters in the production function, the elasticity is not. The relative importance of the inputs in the production process (α, β, and ϕ) is immaterial to the elasticity. As well, the price of input quality, g, does not influence this elasticity. It should not be a surprise that this elasticity is increasing in ρ. As the ability to substitute between inputs increases we naturally expect to see increased substitution of quality for the physical

input in response to increases in the rental rate of the structures required to support the physical input. The elasticity is also increasing in γ.[3] As with the effect of ρ, the higher minimum floor space requirements increase the substitution between physical inputs and the quality of those inputs. Finally, the elasticity is also increasing in the rental rate, r.[4]

Manipulation of equation (2.18) yields conditions under which the elasticity of input quality with respect to the rental rate on structures is elastic. This is the situation that yields the most significant amount of sorting into high rent areas in areas with similar rent gradients. Input quality has an elasticity greater than 1 when $\gamma r > (1 - \rho) w$.

2.1.4 Space per Physical Input

Urban producer theory also generates predictions for a more readily observable quantity; the average space per unit of the physical input. Most commonly this would be the physical space per employee which is a readily tracked number of importance in a wide range of applied analyses including the evaluation of transportation projects (Berechman and Paaswell, 2005), parking requirements (Shoup, 1995), and urban planning models in general (Waddell et al., 2003).

The conditional factor demands provide an expression for the cost minimizing

[3] A proof can be found in Appendix A.

[4] No proof is supplied, however it is identical to the previous result for γ.

amount of space allocated to a physical input.

$$
\begin{aligned}
s &\equiv \frac{S(g,r,w,Q)}{N(g,r,w,Q)} \\
&= \frac{\frac{Q}{A}\left[\left(\frac{\phi}{r}\right)^{\frac{1}{1-\rho}} + \gamma\left(\frac{\alpha}{w+\gamma r}\right)^{\frac{1}{1-\rho}}\right]\left[\alpha\left(\frac{\alpha}{w+\gamma r}\right)^{\frac{\rho}{1-\rho}} + \beta\left(\frac{\beta}{g}\right)^{\frac{\rho}{1-\rho}} + \phi\left(\frac{\phi}{r}\right)^{\frac{\rho}{1-\rho}}\right]^{-\frac{1}{\rho}}}{\frac{Q}{A}\left[\frac{\alpha}{w+\gamma r}\right]^{\frac{1}{1-\rho}}\left[\alpha\left(\frac{\alpha}{w+\gamma r}\right)^{\frac{\rho}{1-\rho}} + \beta\left(\frac{\beta}{g}\right)^{\frac{\rho}{1-\rho}} + \phi\left(\frac{\phi}{r}\right)^{\frac{\rho}{1-\rho}}\right]^{-\frac{1}{\rho}}} \\
s &= \left[\frac{\phi(w+\gamma r)}{r\alpha}\right]^{\frac{1}{1-\rho}} + \gamma
\end{aligned}
\tag{2.19}
$$

The signs of the partial derivatives of this expression are straight-forward except for the effects of r and ρ. The average space per physical input, as one might expect, can be shown to be decreasing in the rental rate of space.

$$
\frac{\partial s}{\partial r} = \left[\frac{1}{\rho-1}\right]\left[\frac{\phi\left(w+\gamma r\right)}{r\alpha}\right]^{\frac{1}{1-\rho}}\left[\frac{w}{w+\gamma r}\right] \; < 0 \; \forall \; \rho \in (-\infty, 1) \tag{2.20}
$$

Whether the average space per physical input is is increasing or decreasing in ρ will depend upon parameter values, as is the case with the effect of ρ on input quality. In this instance it is dependent on the 'productivity adjusted' deployment cost of the physical input relative to that of space.

$$
\frac{\partial s}{\partial \rho} = (1-\rho)^{-2}\left[\frac{\phi(w+\gamma r)}{r\alpha}\right]^{\frac{1}{1-\rho}} \ln\left(\frac{\phi(w+\gamma r)}{r\alpha}\right) \tag{2.21}
$$

While, for the remaining factors I have $\frac{\partial s}{\partial \gamma}, \frac{\partial s}{\partial \phi}, \frac{\partial s}{\partial w} \; > 0$ and $\frac{\partial s}{\partial \alpha} \; < 0$.

Several things are readily apparent. A gradient clearly exists with respect to the rental rate of space and the amount of space allocated to each unit of the physical input. Unlike the input quality gradient, this gradient does not rely on the existence of minimum space requirements. This is not surprising. We expect firms to substitute

the physical input for space (and thereby decrease the space per unit of the physical input) in response to increases in rental rates.

However, the inclusion of minimum space requirements has two effects on space per unit. The minimum space requirement obviously sets a lower bound on the amount of space that can be allocated to physical inputs. In addition, the trade-off between space and physical inputs is also influenced. This can readily be seen by recalling that the deployment cost of a physical input is the direct cost of the physical input plus the space required to deploy the input. An increase in the cost of space increases both the cost of space plus the cost of deploying physical inputs. This has the effect of reducing the change in the relative prices of space and the physical input and thereby reducing the overall effect of changes in space rent on the demand for physical inputs and space.

2.2 The Urban Producer Model with Embodied Input Quality

The embodied form of the urban production function posits a series of physical inputs each of which embodies a different level of quality, as evidenced by its productivity. This framework is quite common in labor economics. Hamermesh (1987) discusses this framework in the context of the demand for labor. Moretti (2011) provides an example of the application of this framework in the modeling of an urban labor market. While the embodied input quality framework may be more commonly observed in the modeling of labor markets than the separable form, as will be seen, it is somewhat less tractable. That said, it also introduces an opportunity to incorporate additional structure in the model.

2.2.1 The Urban Production Function

The urban producer production function based on the embodied form of input quality uses three inputs to produce output: space, S, a high-quality physical input, N_H, and a low-quality physical input, N_L. Each physical input requires a minimum amount of space in order to produce output. The minimum level of space for the high-quality and low-quality physical inputs are γ_H and γ_L, respectively. Formally, the production function is:

$$Q = \begin{cases} A[\alpha_L N_L^{\rho} + \alpha_H N_H^{\rho} + \phi\,(S - \gamma_L N_L - \gamma_H N_H)^{\rho}]^{\frac{1}{\rho}} & \text{if } (S - \gamma_L N_L - \gamma_H N_H) > 0 \\ 0 & \text{otherwise} \end{cases}$$

with

$$\{\alpha_L, \alpha_H, \phi, \gamma_L, \gamma_H\} > 0$$

$$\alpha_L + \alpha_H + \phi = 1$$

$$-\infty \leq \rho < 1$$

As with the urban production function of section 2.1, this function possesses a number of properties that are desirable in a production function. It is homogeneous of degree 1 and consequently homothetic as well. It is also concave in the inputs.[5]

2.2.2 Cost Minimization

Firms face a set of prices for inputs and conditional on the desired level of output, Q, choose the cost-minimizing levels of the inputs. Input prices are r, w_L, and w_H for space, the low-quality physical input, and the high-quality physical input, respectively. Solving the firm's cost minimization problem yields the following set of

[5]No proof of concavity is provided however the proof follows the same form as that in Appendix A.

conditional factor demands:

$$N_L\left(r, w_L, w_H, Q\right) = \frac{Q}{A}\left[\frac{\alpha_L}{w_L + \gamma_L r}\right]^{\frac{1}{1-\rho}} \times \left[\alpha_L\left(\frac{\alpha_L}{w_L + \gamma_L r}\right)^{\frac{\rho}{1-\rho}} + \alpha_H\left(\frac{\alpha_H}{w_H + \gamma_H r}\right)^{\frac{\rho}{1-\rho}} + \phi\left(\frac{\phi}{r}\right)^{\frac{\rho}{1-\rho}}\right]^{-\frac{1}{\rho}} \quad (2.22)$$

$$N_H\left(r, w_L, w_H, Q\right) = \frac{Q}{A}\left[\frac{\alpha_H}{w_H + \gamma_H r}\right]^{\frac{1}{1-\rho}} \times \left[\alpha_L\left(\frac{\alpha_L}{w_L + \gamma_L r}\right)^{\frac{\rho}{1-\rho}} + \alpha_H\left(\frac{\alpha_H}{w_H + \gamma_H r}\right)^{\frac{\rho}{1-\rho}} + \phi\left(\frac{\phi}{r}\right)^{\frac{\rho}{1-\rho}}\right]^{-\frac{1}{\rho}} \quad (2.23)$$

$$S\left(r, w_L, w_H, Q\right) = \frac{Q}{A}\left[\left(\frac{\phi}{r}\right)^{\frac{1}{1-\rho}} + \gamma_L\left(\frac{\alpha_L}{w_L + \gamma_L r}\right)^{\frac{1}{1-\rho}} + \gamma_H\left(\frac{\alpha_H}{w_H + \gamma_H r}\right)^{\frac{1}{1-\rho}}\right] \times \left[\alpha_L\left(\frac{\alpha_L}{w_L + \gamma_L r}\right)^{\frac{\rho}{1-\rho}} + \alpha_H\left(\frac{\alpha_H}{w_H + \gamma_H r}\right)^{\frac{\rho}{1-\rho}} + \phi\left(\frac{\phi}{r}\right)^{\frac{\rho}{1-\rho}}\right]^{-\frac{1}{\rho}} \quad (2.24)$$

The firm's cost function is:

$$C\left(r, w_L, w_H, Q\right) = \frac{Q}{A}\left[r\left(\frac{\phi}{r}\right)^{\frac{1}{1-\rho}} + \left(w_L + \gamma_L r\right)\left(\frac{\alpha_L}{w_L + \gamma_L r}\right)^{\frac{1}{1-\rho}} + \left(w_H + \gamma_H r\right)\left(\frac{\alpha_H}{w_H + \gamma_H r}\right)^{\frac{1}{1-\rho}}\right] \times \left[\alpha_L\left(\frac{\alpha_L}{w_L + \gamma_L r}\right)^{\frac{\rho}{1-\rho}} + \alpha_H\left(\frac{\alpha_H}{w_H + \gamma_H r}\right)^{\frac{\rho}{1-\rho}} + \phi\left(\frac{\phi}{r}\right)^{\frac{\rho}{1-\rho}}\right]^{-\frac{1}{\rho}} \quad (2.25)$$

2.2.3 The Input Quality Gradient

Unlike the separable form of the urban production function, the embedded form does not have an explicit representation for the average quality of inputs employed by the

firm. In this case I calculate the average quality level as the ratio of high-quality physical inputs to low-quality physical inputs. This can be expressed as

$$n\left(r, w_L, w_H, Q\right) \equiv \frac{N_H\left(r, w_L, w_H, Q\right)}{N_L\left(r, w_L, w_H, Q\right)} = \left[\frac{\alpha_H}{\alpha_L}\frac{w_L + \gamma_L r}{w_H + \gamma_H r}\right]^{\frac{1}{1-\rho}} \tag{2.26}$$

The input quality gradient is given by

$$\frac{\partial n}{\partial r} = \frac{1}{1-\rho}\left[\frac{\alpha_H}{\alpha_L}\frac{w_L + \gamma_L r}{w_H + \gamma_H r}\right]^{\frac{1}{1-\rho}}\left[\frac{\gamma_L w_H - \gamma_H w_L}{\left(w_H + \gamma_H r\right)\left(w_L + \gamma_L r\right)}\right] \tag{2.27}$$

The sign of the derivative is equivalent to the sign of $\gamma_L w_H - \gamma_H w_L$ and consequently $\frac{\partial n}{\partial r} > 0$ when $\frac{\gamma_L}{\gamma_H} > \frac{w_L}{w_H}$. The effect depends upon the relative values of exogenous (at least in this model) elements: γ_L, γ_H, w_L, and w_H. While w_L and w_H are exogenous in this demand-side model, they are of course an equilibrium outcome in a more extensive model. Interpreting the physical inputs as high- and low-skilled workers helps provide some context to the relative values of w_L, and w_H. If worker types are largely not an endowment but are based on investments workers make in human capital over their lifetimes then we can assume that $w_H > w_L$ for any worker to be willing to make such an investment. As such, I will assume $w_H > w_L$ for all comparative statics. The minimum floor space requirement is simply a parameter of the production function and little can be said about its true values. The separable urban production function of section 2.1 incorporates a value of γ that does not depend on the quality of the physical input. In the present framework this would be equivalent to assuming $\gamma_L = \gamma_H$. Under this assumption the proportion of high quality inputs employed by the firm is increasing in the rental rate of space.

This necessity of these assumptions illustrates the underlying principle of urban producer theory. Input costs have two components: the price of the actual input as well as the price of the space required to deploy the input. An input with a relative

price significantly above that of another input may still be demanded in levels that exceed that of the lower cost input if the space requirements of the high cost input are low relative to the low cost input.

Comparative Statics of Input Quality

It is readily apparent from observing equations 2.26 and 2.27 that both the input quality level and the input quality gradient are increasing in α_H and decreasing in α_L. This indicates that as the productivity gap between the high and low quality inputs increases we can expect to observe input quality become more responsive to variations in the rental rates of space.

Input quality is clearly decreasing in the minimum space requirement for the high quality input, γ_H. This result is not surprising. As the minimum space that must be allocated to the high quality input increases, the minimum deployment cost, $w_H+\gamma_H r$, increases as well. Similarly, input quality is increasing in the space requirement of the low quality input for the same reasons. The cost-minimizing firm naturally substitutes away from the higher priced inputs and toward lower priced inputs.

The effects of γ_L and γ_H on the input quality gradient are not quite as transparent.

$$\frac{\partial n}{\partial r \partial \gamma_H} = \left(\frac{1}{1-\rho}\right)\left(\frac{\alpha_H}{\alpha_L}\right)^{\frac{1}{1-\rho}} (w_L+\gamma_L r)^{\frac{\rho}{1-\rho}} \times \left[\frac{\rho-2}{1-\rho} r \left(w_H+\gamma_H r\right)^{\frac{2\rho-3}{1-\rho}} \left(\gamma_L w_H - \gamma_H w_L\right) - w_L \left(w_H+\gamma_H r\right)^{\frac{\rho-2}{1-\rho}}\right] \quad (2.28)$$

Under the maintained assumption that $\frac{\gamma_L}{\gamma_H} > \frac{w_L}{w_H}$, both terms in the brackets are negative for admissible values of ρ. Therefore the input quality gradient is decreasing in γ_H. The input quality gradient is increasing in γ_L under the same assumption.[6]

The effects of ρ on input quality as well as the input quality gradient are ambigu-

[6] A proof can be found in appendix A

ous. Even with the assumptions regarding γ_L, γ_H, w_L, and w_H it is not possible to put a sign on the derivatives.

$$\frac{\partial n}{\partial \rho} = (1-\rho)^{-2}\left[\frac{\alpha_H}{\alpha_L}\frac{w_L + r\gamma_L}{w_H + r\gamma_H}\right]^{\frac{1}{1-\rho}} \ln\left(\frac{\alpha_H}{\alpha_L}\frac{w_L + r\gamma_L}{w_H + r\gamma_H}\right) \tag{2.29}$$

As can be seen, the standard assumptions are unable to restrict the term of the natural log to being consistently greater or less than 1 and therefore the sign of the term is indeterminate. The conditions that will lead to a positive effect mirror those of the separable urban production function. When the productivity of the high quality input, as measured by α_L, is high relative to that of the low quality input and its deployment cost is low relative to that of the low quality input then the firm will substitute toward the high quality input in response to increases in the cost of space. The expression for $\frac{\partial n}{\partial r \partial \rho}$ is not as clean; however the same issue arises.

Elasticity of Input Quality

The elasticity of input quality with respect to the rental rate of structures can also be calculated.

$$\epsilon_{nr} \equiv \frac{\partial \ln n}{\partial \ln r} = \frac{r}{1-\rho}\left[\frac{\gamma_L w_H - \gamma_H w_L}{(w_L + \gamma_L r)(w_H + \gamma_H r)}\right] \tag{2.30}$$

Of course, the sign of the elasticity, ϵ_{nr}, is equivalent to the sign of the input quality gradient itself. It is clear that the elasticity is increasing in ρ, as in the separable model. However, unlike the separable model, few definitive statement with regards to the effect of r on the input quality elasticity can be made without additional assumptions. The previous restrictions on the values of these parameters in this partial equilibrium model do not help resolve the effect. However, it will be the case

that the elasticity is increasing in r if $w_L w_H - \gamma_L \gamma_H > 0$, in addition to the earlier condition that $\gamma_L w_H - \gamma_H w_L > 0$.[7] Basically, the minimum space requirements cannot be too large.

As with input quality and its gradient, the input quality elasticity is decreasing in γ_H and increasing in γ_L under the maintained assumption that $\gamma_L w_H - \gamma_H w_L > 0$.[8] Finally, as with the separable model, it is clear that the elasticity is increasing in ρ. As described earlier, as the ease of substituting between inputs increases, we expect to see factor demands which are much more responsive to prices.

2.2.4 Space per Physical Unit of Input

The space per physical unit of input can also be examined. Unlike the separable model, in the embodied model there are two physical inputs: low quality and high quality. Because space is undifferentiated in the model, it is not possible to separate space devoted to the low quality input from space devoted to the high quality input. However the average space allocated to all physical units of the input can be examined.

$$\begin{aligned} s &\equiv \frac{S(r, w_L, w_H, Q)}{N_L(r, w_L, w_H, Q) + N_H(r, w_L, w_H, Q)} \\ &= \frac{\left(\frac{\phi}{r}\right)^{\frac{1}{1-\rho}} + \gamma_L \left(\frac{\alpha_L}{w_L + \gamma_L r}\right)^{\frac{1}{1-\rho}} + \gamma_H \left(\frac{\alpha_H}{w_H + \gamma_H r}\right)^{\frac{1}{1-\rho}}}{\left(\frac{\alpha_L}{w_L + \gamma_L r}\right)^{\frac{1}{1-\rho}} + \left(\frac{\alpha_H}{w_H + \gamma_H r}\right)^{\frac{1}{1-\rho}}} \end{aligned} \tag{2.31}$$

[7] This result can be found in appendix A

[8] The former result is obvious and the latter result can be found in appendix A

This expression is relatively complex, however by implementing the assumption that $\gamma_L = \gamma_H$, this cleans up to a more manageable expression.

$$\widetilde{s} = \frac{\left(\frac{\phi}{r}\right)^{\frac{1}{1-\rho}}}{\left(\frac{\alpha_L}{w_L+\gamma r}\right)^{\frac{1}{1-\rho}} + \left(\frac{\alpha_H}{w_H+\gamma r}\right)^{\frac{1}{1-\rho}}} + \gamma \tag{2.32}$$

The signs of most of the partial derivatives are readily apparent by casual observation. The space per physical unit of input under this special assumption is increasing in γ, ϕ, w_L, and w_H. It is decreasing in α_L, α_H, and r.[9]

2.3 Summary

This chapter has presented two alternative formulations for incorporating input quality into a production function that exhibits congestion. The implications of these two models in an urban landscape are similar. This indicates that it is not the convention for modeling input quality that drives the results. An input quality gradient is shown to exist. The input quality gradient describes the positive correlation between the rental rate of structures and the quality of inputs deployed by the firm.

Analysis of each of the models has helped identify the situations in which the input quality gradient is likely to be the most pronounced. Equation (2.15) and figure 2.2 demonstrate that production processes in which the relative importance of the physical input, such as labor, is less important than it's quality component, human capital, are likely to have a steeper input quality gradient. This result provides guidance in selecting an industry in which to test for the implications of urban producer theory. In the following chapter I test for the existence of an input quality gradient in the Cleveland CBSA by examining how the quality of attorneys varies with the

[9]The proof of this final result is in appendix A.

rent at their places of employment. The chapter is then followed by a discussion of the implications of the input quality gradient in a number of settings.

Chapter 3

Testing Urban Producer Theory

In this chapter I test the primary implication of urban producer theory. The theory hypothesizes that worker quality is an increasing function of the rental rate of space as firms substitute labor quality for labor quantity in response to higher floor area rents. Two basic pieces of information are required in order to test the theory with respect to labor inputs: 1) employee human capital and 2) floor area rent of the employment location. Several sources are used to get this information to allow for the testing of urban producer theory on attorneys in Ohio. Attorneys provide an opportunity to test for the labor quality gradient as predicted by urban producer theory. Attorneys are observationally equivalent by a commonly used measure of human capital; education. However, there is substantial evidence including formal studies ranking law schools, that degrees from some schools are considered to be of a higher quality than degrees from other schools.[1] This provides a prime opportunity to test for an input quality gradient on a labor force that traditional human capital measures would fail to fully control for the unobserved heterogeneity. In addition, as demonstrated in chapter 2, the input quality gradient is steeper when human capital is more important than

[1]Section 3.1.1 discusses previous uses of law school ranks as a measure of unobserved ability in attorneys.

physical labor in the production process. Clearly human capital is more important than physical labor in the production of legal services. However, space is also an important input, as attorneys require offices in which to perform their work and therefore the minimum space requirement is likely to be substantially larger than an industry such as retailing. Finally, as illustrated by figure 3.2, attorneys are widely distributed across the Cleveland metropolitan area in keeping with a basic assumption regarding urban producer theory. These factors will increase the likelihood that an input quality gradient for attorneys can be found. Of course, there are also factors associated with attorneys that make it less likely that an input quality gradient will be found. If high quality attorneys demand more space than low quality attorneys, it may be the case that the deployment cost of high quality attorneys relative to low quality attorneys is such that the input quality gradient is reversed.[2] Furthermore, if law firms are not well-managed or do not strictly minimize costs, I am less likely to find evidence of an input quality gradient. Finally, sorting of attorneys based on quality and school location may limit my ability to measure the input quality gradient. As demonstrated by table 3.2, none of the top law schools are located in Ohio. If the highest quality attorneys choose to migrate to work locations in the largest metropolitan areas such as New York and Chicago, then the graduates from the top law schools that are employed in Cleveland are of a lower quality than their classmates located in New York or Chicago. This inter-metropolitan sorting of lawyers can lead to a situation where school rankings on a local level are more compressed than are rankings on the national level. This will make it more difficult to find an input quality gradient.

The chapter begins by introducing the data used to determine attorney quality

[2]The embodied quality model of section 2.2 allows for such a situation and shows that the input quality gradient will be positive when the minimum floor area requirements of the high quality input are sufficiently greater than the minimum floor area requirements of the low quality input.

and the rent at the attorneys' places of employment. This is followed by a discussion of the estimation strategy used to test for the labor quality gradient. The chapter concludes with a series of tests for the existence of a labor quality gradient in the Cleveland CBSA.

3.1 The Data

I use two main data sources to conduct the empirical investigation. Detailed characteristics of attorneys admitted to the Ohio bar are used to determine the human capital of workers at a location. Data from CoStar Group on available leases provides information on the rental rate of space in the Cleveland CBSA. I discuss each data source in turn.

3.1.1 Attorney Data

Rules governing the Ohio bar require attorneys admitted to the practice of law in Ohio to register with the Court. A large portion of the registration information is a public record and therefore readily available to researchers. I have obtained the records for all active attorneys in Ohio from the Office of Attorney Services of the Ohio State Supreme Court. There are 35,664 active attorneys as of August 1, 2009 that are members of the Ohio bar. The data contains information on the attorney's employer as well as some personal information on the attorney. The records contain the name, address, and phone number of the attorney's employer. Personal information on the attorney includes name, date of birth, law school attended, job title, and gender. Information on the attorney's residence is not part of the public record. Additional information in the record includes when and how the attorney was admitted to the Ohio bar, as well as a unique registration number.

The information on the law school attended by the attorney is combined with US News and World Report's ranking of US law schools. I assume that attorneys graduating from higher ranked law schools will have higher levels of human capital. Use of law school quality as a measure of the human capital possessed by attorneys is far from a novel idea. Garicano and Hubbard (2005) use US News law school rankings to proxy for the "cognitive ability" of attorneys. Spurr (1990) and Laband and Lentz (1993) use measures of law school quality as a measure of attorney quality in order to control for worker heterogeneity when testing for evidence of sex discrimination in the legal profession. Boylan and Long (2005) measure the quality of federal prosecutors based on whether they attended elite law schools. Tam (1997) analyzes the wage premium that accrues to graduates of highly ranked law schools and identifies problem-solving skills as a significant contributor to the premium.

US News and World Report ranks 184 law schools. The 102 highest ranked schools are ordered, though there are some ties. The remaining 82 schools are sorted into two ordered tiers, tiers three and four, however they are not ordered within each tier. US News provides several additional pieces of information on these schools such as attendance, admission rates, tuition, employment rates at graduation, and starting salaries.

Refining the Attorney Sample

I refine the list of the 35,664 active attorneys of the Ohio bar for the analysis. My first refinement is to classify attorneys according to the type of organization with which they are employed. Based upon the attorney's job title and employer name, they are classified as being in the private, government, or academic sectors. Unemployed attorneys were not classified. The data contain 27,612 private sector attorneys, 6,243 attorneys employed by governmental organizations, and 491 attorneys associated with

academic institutions. The remaining 1,018 attorneys are unclassified because they are unemployed or otherwise did not list an employer, or work for an organization that I could not identify.

The next step of the refinement is to harmonize employer names of the private sector attorneys to allow a calculation of firm size within Ohio and within the Cleveland CBSA. One feature of the data is that they are self-reported, commonly by individual attorneys, and there is scant evidence that the Office of Attorney Services undertakes extensive data quality controls. This leads to substantial variation in the exact form that attorneys report for their titles, employer names, and employer addresses. Attorneys who work for the same employer at the same location may list slight variations in the name and address of their employer. Correcting employer names is a manual procedure in which I examine similarly spelled employer names to determine if they are the same firm based on the reported employer address or based upon the firm's website listing additional office locations.

Once employer names are corrected, I focus on the 8,547 private sector attorneys that the Ohio bar identifies as being employed in the counties that comprise the Cleveland CBSA. Cleveland is chosen for the analysis because it is the largest city in Ohio and is likely to have the most significant rent gradient. Furthermore, unlike Cincinnati, the metropolitan area is wholly within the state of Ohio, ensuring that all attorneys in the area are observed. Finally, Cleveland has the largest number of attorneys of any metropolitan area in Ohio. I only analyze private sector attorneys because urban producer theory is based upon the cost-minimizing behavior of firms and so I select the employers which are most likely to respond to differences in floor area rents. Attorneys that did not graduate from a law school ranked by US News were also excluded from analysis. Most of these law schools consisted of foreign law schools such as the Universities of Gdansk and Manitoba as well as law schools that

are no longer in existence such as the Antioch School of Law in Ohio. This culling further reduces the number of attorneys available for analysis to 8,501.

Once the sample is refined, the street addresses of the employers of the 8,501 attorneys are assigned a latitude and longitude using the WebGIS geo-coding application available online from the GIS Research Laboratory at University of Southern California.[3] The geo-coding process provides a latitude and longitude at the finest level of geography that can be identified by the software based on the address information supplied. The finest level of geographic detail possible is the street address level which indicates that the software is able to identify the segment of the street that contains the numeric address on the attorney registration form. The software reaches this level of precision by locating the numeric address along the named street that is within the ZIP code provided on the registration form. Failure to find a precise match can arise for several reasons. If the ZIP code is wrong, then the WebGIS software will be unlikely to find a matching street name and numeric address in the incorrect ZIP code. Other sources of error occur if the numeric address provided is not assigned to a street segment either because of an error on the registration form or because the numeric address did not exist at the time the geographic database used by WebGIS was created.[4] In instances where an accurate match cannot be found, the software indicates that the match quality is at the ZIP code level or in cases where the ZIP code on the registration form does not exist, an even higher level of geography such as the latitude and longitude of the city center is provided. The registration data is run through the WebGIS software as a preliminary matter. Records that are not located at the street address level are then individually examined. I search for these

[3]DW Goldberg and JP Wilson, USC WebGIS Services. Available online at `https://webgis.usc.edu`.

[4]The main source of location information used by WebGIS are the 2008 US Census Bureau TIGER/Line files.

addresses using Google's online mapping tool "Google Maps."[5] If I find a match using this product, I attempt to determine the cause of the problem. A common problem is consistent spelling. For example, Public Square is a major road in the heart of the Cleveland CBD. The WebGIS application does not recognize abbreviations such as "Public Sq." Therefore I edit these records for street name consistency. Additional problems occur when the numeric addresses or streets are not included in the WebGIS database due to new construction. In these cases I either adjust the numeric address to one that exists in the WebGIS database, or in cases where the entire street is newly constructed, I use a nearby street and numeric address. These cases are mostly found in suburban locations. The benefit of using Google Maps for this process is twofold. Its database is constantly updated for construction and in addition it commonly shows the location of the firm that employs the attorney. Using this process, I obtain a street level precision for 8,286 private sector attorneys in the Cleveland CBSA that graduated from a law school ranked by US News. Therefore 97.5% of the private sector, US-educated attorneys in the Cleveland CBSA are successfully geo-coded.[6]

An additional benefit of the geocoding process is that the geocoder returns a matched address for those records it successfully locates. This yields a consistent address for records which may originally have slight variations in the street address of the employer.

The final selection criterion is to limit the analysis to attorneys that have complete

[5] `http://maps.google.com/`

[6] The rate at which attorneys were successfully geocoded does vary based on the rank of the law school. I use a binary probit model to examine the correlation between successful geocoding and law school rank. The results indicate a statistically significant relationship between the probability of being successfully geocoded and the school rank. Attorneys that attended highly ranked law schools were more likely to be successfully geocoded. The probit model predicts that the top ranked attorneys had a probability of success of 98.9% and the lowest ranked attorneys had a probability of success of 96.1%. In the data, 100% of the top ranked attorneys were successfully geocoded while only 96.3% of the lowest ranked attorneys were successfully geocoded.

data on bar admission date, birth date, and gender in order to have a single sample of complete cases for all analysis. This yields a final sample of 8,283 private sector attorneys employed at 1,981 locations in the Cleveland CBSA.

Descriptive Statistics of the Attorney Data

The distribution of firm and employer size within the Cleveland CBSA is shown in figure 3.1. The distribution of firm sizes in the data indicate that 77% of the firms employ a single attorney while 91% employ less than 4 attorneys. The distribution of employer sizes indicates that 32% of private sector attorneys in the Cleveland CBSA work for an employer with a single attorney in the CBSA and 45% work for an employer with fewer than 4 attorneys. Table 3.1 lists the largest private sector employers of attorneys in the Cleveland CBSA. 21% of the private sector attorneys in the Cleveland CBSA work for these 18 firms.

Table 3.2 lists the top-10 ranked law schools according to US News. In addition, the table lists the number of graduates from the law school in the attorney data. 6.5% of the private sector attorneys in the Cleveland CBSA graduated from one of these law schools.

Table 3.3 shows the law schools that are most commonly found in the sample designated for analysis. Of particular note is that over 60% of the private sector attorneys in the Cleveland CBSA graduated from a law school located in the CBSA.

I specify 127 Public Square as the center of the Cleveland CBD. With 57 floors, this building, also known as Key Bank Tower, is the tallest building in Ohio and the 18th tallest in the United States. Table 3.4 shows the density of attorneys in the sample in the Cleveland CBSA. A clear attorney density gradient exists. Figure 3.2 shows the locations of the private sector attorneys in the Cleveland CBD.

Descriptive statistics of the other relevant data in the attorney dataset is in table

3.5. The ages of active members of the Ohio bar is 25 to 96 years of age while their experience levels, as measured by the years since being admitted to the Ohio Bar, ranges from 0 to 70. The median age is 49 and the median experience level is 20 years. 73% of the active Ohio Bar is male.

3.1.2 Rent Data

I use information from CoStar Group's commercial property database. The CoStar Group claims to be commercial real estate's leading provider of information, analytic and marketing services. It conducts expansive, ongoing research to produce and maintain the largest and most comprehensive database of commercial real estate information.[7] The database contains information on commercial properties including leasing and sales information. The CoStar database has been used by a number of researchers in the urban economics and real estate fields. Among the recent research using the CoStar data are Fuerst and McAllister (2011), Eichholtz et al. (2010), Wheaton et al. (2009), and Abadie and Dermisi (2008).

Refining the Sample of Rental Data

CoStar provided me with an extract of all properties in the Cleveland CBSA that are in their database. In addition, leasing information was provided for those properties that had space available for lease as of February 2011.[8] The data were provided in two tables. The property table contains information on 3,822 buildings in the CoStar database that are located in the Cleveland CBSA. Table 3.6 lists the information available for each of the buildings in the property table. The space table contains information on 4,040 individual spaces offered for lease in the Cleveland CBSA in-

[7] See `http://www.costar.com/about/`

[8] These spaces are 'currently marketed' in the terminology used in the real estate industry.

cluding the asking rent and other characteristics of the space. Table 3.7 lists the fields available for each record in the space table. Not all properties in the data have corresponding space records. Some properties simply did not have space for lease at the time the data was extracted. The space records in the CoStar data are associated with 1,080 different buildings in the Cleveland CBSA.

All records in the property table, regardless of whether space is for lease in the building, are run through the WebGIS geo-coding application from the GIS Research Laboratory at the University of Southern California. Similar procedures to those applied to the attorney data are used to correct street addresses that are not matched by the geocoder. Of the 3,822 properties, 3,778 are located at the street address level. These properties contain 3,974, or 98.4% of the 4,040 spaces offered for lease in the data.

As with the attorney data, complete cases of the rental data are selected for analysis. Complete cases are those which were successfully geocoded at the street address level. In addition, they have complete information on the asking rent, services included in the lease, building size (square footage and stories), building class, and building age. The sample is restricted to spaces that are offered for lease strictly as office space, as opposed to retail, medical, or some combination of uses. The final refinement is to exclude spaces that are in the lower level or basement of a property.[9] Even though these spaces are available as office space, I generally assume that these are not the types of spaces that attorneys would choose. This final yield is 2,849 spaces in 752 properties.

In addition to having rental rates for individual spaces, I develop an average asking rent and space characteristics for a property as well. A weighted average of the asking rent and characteristics of spaces in a property is calculated as well using the refined

[9] 117 office spaces in the data are in the basement or lower level.

space dataset.[10] The weights consist of the area of the space offered for lease as a fraction of the total area of all spaces offered for lease in the building.

Descriptive Statistics of the Rental Data

Table 3.8 presents descriptive statistics of the properties in the database. Statistics for all properties in the database as well as statistics for properties containing the office spaces for lease that are used in later analyses are presented. It appears that buildings with an office space available for lease are different than the remaining properties. This is illustrated more formally in table 3.10 which shows the results of a binary logit model that examines the influence of the characteristics of a property on the the probability that the property contains an office space available for lease that is used in the analysis. As can be seen, properties with available space tend to be larger, newer, and higher quality. This finding is not terribly surprising. Properties with a larger number of office spaces are more likely to have a vacancy at a given point in time than are properties with only a handful of office spaces regardless of whether building size has an influence on the desirability of a property. A similar effect may be at work for building age as well. A building begins its lifecycle with 100% vacancy rate and attracts tenants during its construction phase. Properties that fail to lease fully prior to the completion of construction may be the cause of the finding that newer properties are more likely to have space available for lease.

Figure 3.3 shows the locations of the properties in the Cleveland CBD and distinguishes those with and without office space for lease. The influence of transportation networks on the location of commercial properties is apparent. For example, the string of properties leading south along the western side of the CBD are following

[10]Characteristics that may vary among spaces in the same property are the asking rent, the services included in the lease, the size of the space and its location within the building. While the permitted use of the space may also vary, I restrict my analysis to office spaces.

I-71 while the linear pattern paralleling Lake Erie to the north and east of the CBD represents the I-90 corridor.

Table 3.9 presents descriptive statistics from the space data used in the analyses that follow. The size of the spaces range from 55 square feet up to 95,000 square feet. The asking rent on these spaces ranges from $1.10 per square foot per year to $45.00 per square foot per year. Figure 3.4 illustrates the density of asking rents for spaces and demonstrates that very few spaces rent for more than $30.00 per square foot per year. Slightly more than 40% of the spaces are offered at full gross rent which includes utilities and cleaning services. A slightly greater proportion of spaces do not include all of these services in the rent for the space. Finally about 10% of spaces are being offered on net leases, either single, double, or triple net. These leases do not include any services and in addition require the lessee to cover property maintenance, property taxes, and possibly other costs generally assigned to a property owner.

3.2 Estimation Strategy

Urban producer theory posits that the quality of an employee is increasing in the rent of a location. I use an ordered response model to estimate the correlation between attorney quality, as measured by the law school attended by the attorney, and the rent at the attorney's place of employment. I use an ordered response model because the law school rankings do not represent cardinal measures of quality but rather indicate where schools lie in an ordering of the quality. For example, a 25% increase in the rank of a law school does not indicate that a graduate from the higher ranked school is 25% more productive than a graduate from the lower ranked school. The disadvantage of this framework is that estimation requires strong distributional assumptions. Throughout the analysis I assume that the error terms of the underlying

latent variable have a standard logistic distribution. I choose the logistic assumption, rather than the probit, because of the ease of calculating odds ratios and other summary statistics regarding the fitted model.[11]

The urban producer hypothesis is tested with an ordered logit model. The model estimates the expectation of the quality of the law school that the attorney attended conditional on the office rent where the attorney is employed. Formally, the underlying latent variable of the ordered logit model relates law school quality, H^*, a continuous though unobserved variable, to the office rent, r.

$$H_i^* = \delta + \beta r_i + \mu_i \tag{3.1}$$

This latent variable model generates the following ordered outcomes that translate the continuous law school quality outcome, H^*, to a ranking of the law school quality, H. The data generating process is

$$H_i = \begin{cases} 1 & \text{if } H_i^* \leq \alpha_1 \\ 2 & \text{if } \alpha_1 < H_i^* \leq \alpha_2 \\ & \vdots \\ B-1 & \text{if } \alpha_{B-2} < H_i^* \leq \alpha_{B-1} \\ B & \text{if } \alpha_{B-1} < H_i^* \end{cases} \tag{3.2}$$

[11] As Greene and Hensher (2010, page 138) point out '... one of the ambiguities in the set of procedures for ordered choice modeling is the distributional assumption. There seems to be little to determine whether the logit, probit, or some other distribution is to be preferred.' Sensitivity tests indicate that the overall qualitative results of my analyses do not depend on this distributional assumption.

The log-likelihood function used to estimate this ordered logit model is[12]

$$
\begin{aligned}
L_2(\alpha,\beta) = \sum_{i=1}^{N} \Bigg\{ & m_{i1} \ln\left[\frac{1}{1+e^{\beta r_i-\alpha_1}}\right] \\
& + \sum_{b=2}^{B-1} m_{ib} \ln\left[\frac{1}{1+e^{\beta r_i-\alpha_b}} - \frac{1}{1+e^{\beta r_i-\alpha_{b-1}}}\right] \\
& + m_{iB} \ln\left[1 - \frac{1}{1+e^{\beta r_i-\alpha_{B-1}}}\right]\Bigg\}
\end{aligned}
\tag{3.3}
$$

where $m_{ib} = 1$ when $H_i = b$ and 0 otherwise.

I use three estimation strategies to test the hypothesis. The first strategy is to eschew the use of rent data and use the distance from the CBD as a proxy for rent. This analysis relies on the well-documented inverse relationship between the distance from the CBD and land prices.[13] A second strategy is based on matching buildings housing attorneys in the attorney data with buildings in the CoStar property data and using the rent of the matched buildings to test urban producer theory. My final strategy implements a two-step procedure which estimates the rent at each location of an attorney in the data before testing urban producer theory on the full sample of attorneys.

3.2.1 Evaluating the Estimates from the Ordered Logit.

The underlying latent variable, H^* of the ordered logit model is an abstract concept with little inherent value. It does not relate in any quantitative way to the input quality of section 2.1, nor does it indicate any cardinal productivity measure. As such, the marginal effect of space rent on the latent variable does not provide a measure of the magnitude, or economic significance, of the estimated relationship. In

[12] Implicit in the log-likelihood function is the scaling of the coefficients by the variance which is set at $\frac{\pi^2}{3}$.

[13] Haughwout et al. (2008) provides a recent example of the land rent gradient in New York.

order to obtain such a measure, I focus on the probabilities of attorney quality levels. In particular, I measure the effect by examining the influence of office rent on the probability of the highest and the lowest quality attorneys working at a particular location. I choose the highest and lowest quality levels because, as is evident from equation (3.2), it is these levels where the direction of the marginal effect on the probability is constant based on the sign on the estimated coefficient on space rent; the probability of the highest quality attorney is increasing when the estimated coefficient on rent is positive and decreasing when the estimated coefficient is negative, and conversely for the lowest quality level. The direction of the marginal effect at the other quality levels will depend upon the values of the so-called ancillary parameters, α_b.

I present two calculations to examine the magnitude of the effect; generally in the form of a figure to illustrate their values over the observed range of office rents. The estimated probabilities of observing a high and a low quality attorney at a particular location based on the space rent at the location are calculated. Urban producer theory predicts that the probability of observing a high quality attorney is increasing in the space rent, while the probability of observing a low quality attorney is decreasing. The probabilities, derived from the logistic distribution can be seen in equation (3.3). They are

$$P(\text{Low Quality} \mid \text{Rent}) = \left[1 + e^{\widehat{\beta}\text{Rent} - \widehat{\alpha}_1}\right]^{-1}$$
$$P(\text{High Quality} \mid \text{Rent}) = 1 - \left[1 + e^{\widehat{\beta}\text{Rent} - \widehat{\alpha}_{B-1}}\right]^{-1}$$

In addition to examining the effect of rent on the probability of high and low quality attorneys, I also examine how responsive these probabilities are to changes in space rents. I do this by calculating an elasticity of the probabilities with respect to

rent.

$$\frac{\partial \ln P(\text{Low Quality} \mid \text{Rent})}{\partial \ln \text{Rent}} = -\widehat{\beta}\text{Rent}\left[1 + e^{\widehat{\beta}\text{Rent}-\widehat{\alpha}_1}\right]^{-1} e^{\widehat{\beta}\text{Rent}-\widehat{\alpha}_1}$$

$$\frac{\partial \ln P(\text{High Quality} \mid \text{Rent})}{\partial \ln \text{Rent}} = \widehat{\beta}\text{Rent}\frac{\left[1 + e^{\widehat{\beta}\text{Rent}-\widehat{\alpha}_{B-1}}\right]^{-2} e^{\widehat{\beta}\text{Rent}-\widehat{\alpha}_{B-1}}}{1 - \left[1 + e^{\widehat{\beta}\text{Rent}-\widehat{\alpha}_{B-1}}\right]^{-1}}$$

It is the elasticity which may be the most relevant to judge the economic significance of the input quality gradient for attorneys in the Cleveland CBSA. It provides the most direct measure of the responsiveness of firms to changing rental rates for space.

3.3 Estimation Based on Distance from the Central Business District

I begin with a less formal model of the spatial-quality relationship. Rather than analyzing the relationship between space rents, which may be measured with some amount of error, I begin with an analysis of the spatial location of employment as it relates to input quality. In the urban environment, the key component of a commercial location is the distance to the CBD. This distance is related to the price of land through the rent gradient. The rent gradient is a theoretical outcome of the standard monocentric urban model such as that described by Brueckner (1987). It's existence, both as a pure land rent gradient and as a gradient associated with developed properties, has been verified empirically by a diverse group of studies including home prices by Coulson (1991), vacant land by Haughwout et al. (2008), apartment building prices by Söderberg and Janssen (2001), and office rents by Sivitanidou (1995).

I find that a small rent gradient for office space exists in the Cleveland CBSA. Table 3.11 presents the results of a log-linear regression using the CoStar data. The natural log of the asking rent of an office space is regressed on the natural log of the miles from the Cleveland CBD as well as a set of variables to control for the quality of the office space.

Given the existence of an office rent gradient in Cleveland, a test of urban producer theory would consist of testing for a negative correlation between the quality of an attorney and the distance of the place of employment from the CBD. The results of the estimation of this distance model are shown in table 3.12. The results are supportive of the major implication of urban producer theory. High quality attorneys are more likely to be located closer to the CBD than are low quality attorneys. However, as discussed in section 3.2, the estimation results in table 3.12 do not provide an indication of the magnitude of the effect. I examine the effect of distance from the CBD on the probabilities of observing high quality and low quality attorneys at a location in order to assess the economic significance of the finding. Figure 3.5 illustrates the effect of distance on the probabilities and elasticities of low and high quality attorneys locating in a property based on the distance from the CBD. It is immediately apparent that low quality attorneys are always more prevalent because they dominate the population of attorneys in Cleveland. As seen in table 3.2, there are only 23 of the highest quality attorneys in Cleveland. This compares to 309 of the lowest quality attorneys. Given the concentration of attorneys close to the CBD, a very strong sorting mechanism based on distance to the CBD would have to be at work for the probability of the highest quality attorney to exceed that of the lowest quality attorney. However, the figure does show evidence of sorting at work. The probability of observing the lowest quality attorney rises from 2.7% at the CBD to 13.2% when 40 miles form the CBD. This nearly five-fold increase mirrors a similar

decline in the probability of observing the highest quality attorney at those distances.

3.4 Estimation Based on Matched Locations

The second strategy I employ for identifying the rent at an attorney's place of work is to match the employer addresses to records in the CoStar data. The use of the WebGIS geocoding service makes this a relatively painless operation. The service returns a standardized street address for each record that is successfully geocoded and therefore matching is not contingent on correcting for subtle variations in the naming conventions of addresses. I am able to successfully match 4,072 of the 8,283 attorneys in the sample.[14] While this is a substantial fraction of the attorney dataset, it is a much lower fraction of the locations where attorneys are employed. There are 1,981 unique locations where attorneys are employed and I am able to match up 279 of those for a success rate of 14%. The matched properties have asking rents ranging from \$6.44 to \$45.45 per square foot per year with a mean of \$18.72 and a standard deviation of 5. This will ensure that I have sufficient variation in rents in order to identify the input quality gradient if it exists.

Table 3.13 presents the results of a logit regression on the match success. Certain types of attorneys are more likely to work in a building in the CoStar database. Higher quality, younger, yet more experienced, attorneys working for smaller firms located closer to the CBD are more likely to be matched up with a CoStar property.

[14] 20 attorneys at 6 locations were matched up with a property address that was listed twice in the property database. An additional 4 attorneys at 3 locations matched up with 3 property addresses in the property database. Rather than exclude the duplicated properties, the attorney is included in the estimation data with a record for each of the matched properties.

3.4.1 Results of Matched Property Analysis

Table 3.14 presents the results of the estimation of the ordered logit model to predict attorney quality as a function of average asking rent in the building housing the attorney. The results do not reject the basic proposition of urban producer theory. Higher quality attorneys are located in buildings with higher rents. Figure 3.6 illustrates the effect of office rent on the probabilities and elasticities of low and high quality attorneys being employed at a location. Of interest is that the elasticity is increasing in the rent level. This is one of the implications of the model presented in chapter 2.1.

While model 1 does provide evidence in support of urban producer theory, the existence of an office rent gradient in the Cleveland CBSA, could lead to an omitted variable problem. The rent gradient estimated in table 3.11 indicates a correlation between distance from the CBD and floor area rents. Failure to include distance from the CBD in the estimation can lead to omitted variable bias that may generate false conclusions regarding the effect of office rents on attorney quality. In order to eliminate this potential bias, table 3.14 also presents estimation results from the ordered logit model that includes distance from the CBD as an explanatory variable as well as the rent in the attorney's building. The effect of distance from the CBD on the quality of an attorney cannot be distinguished from zero at the 99% confidence level, while the floor area rent continues to have a statistically significant effect on the quality of attorneys located in the building. Figure 3.7 illustrates the predicted elasticities and probabilities of low and high quality attorneys based on the rent at a location. These predictions are calculated at the median and mean values of the attorneys' distance from the CBD, 0.5 miles and 6.4 miles, respectively.

An additional factor that may influence the locations at which attorneys are employed may be the size of the firm. Table 3.1 lists the largest employers of attorneys

in the Cleveland CBSA. 92% of the attorneys working for these firms are located within one-half mile of the CBD. The data exhibit a strong negative correlation between employer size and distance to the CBD. Attorneys working for large firms are more likely to be located in offices closer to the CBD. This finding mirrors that of Limehouse and McCormick (2011). One concern in regards to models 1 and 2 is that the failure to account for firm size may drive the reported results. Firm size and input quality may well be correlated, possibly because the higher quality inputs are more productive and thereby contribute to the success of the firm or because larger firms signal the quality of the output by employing higher quality inputs and locating in high rent buildings. An additional concern may be that large firms choose to locate centrally, where rents happen to be high, in order to have easier access to a greater number of potential clients.

There is a significantly greater amount of work which offers explanations for firms' locational choices on an inter-urban or regional scale. Possibly most relevant to the discussion at hand is a firm's decision to locate in a cluster of similar firms such as the high density of information technology companies in California's Silicon Valley. Almazan et al. (2007), expanding on earlier work, develop a model of firm behavior that suggests that the more efficient firms locate in clusters while the less efficient operate in isolation. On an empirical basis Andersson et al. (2007) find that higher quality workers are more likely to work for efficient firms and that this effect is enhanced by the density of the local labor market. Similarly, Mion and Naticchioni (2009) find a positive correlation between worker quality and firm size in Italy. Consequently, given the correlation between rents and distance from the CBD and the corresponding correlations between distance from the CBD, firm size, and worker quality, a legitimate concern arises over a potential positive bias in the coefficient for testing urban producer theory.

Regardless of the mechanism generating the correlation, I also estimate the ordered logit model using the number of attorneys employed by the firm as a control. Models 3 and 4 in table 3.14 presents the results. The rent at a location continues to have a statistically significant effect on the quality of an attorney, as does the size of the employer. Figures 3.8 through 3.10 illustrate the elasticities and probabilities for models 3 and 4.

Direct comparison of the estimated coefficients for rent of the four models is inappropriate. The cutpoints, α_b, are an intrinsic portion of the model and they will also change between models. Therefore I compare the elasticities and probabilities of observing low and high quality attorneys for each of the two models. In addition, because the probabilities are dependent not only on rent but the distance from the CBD and employer size, the calculations are conditioned on the median value of the distance from the CBD, 0.5 miles, and the median employer size, 5 attorneys. Figure 3.11 shows the influence of rent on the elasticity and probability of high and low quality attorneys for the four models. The immediate take-away from this comparison is the minor role that distance from the CBD plays in explaining the presence of high and low quality attorneys. Adding the distance measure to model 1, as is done with model 2, has very little effect on the predicted values. Similarly, the predicted values of model 3 and 4 are nearly indistinguishable.

3.5 Estimation Based on Predicted Rents

My final strategy estimates the rent at each employment location and thereby allows use of the full attorney sample. There are several possible methods to use in estimating the rent at a location. One naive method would be to simply use the estimated office rent gradient in the Cleveland CBSA from table 3.11 and then use the distance

from the CBD of the attorney locations to test the urban producer hypothesis. There is very little difference between this and the simple model of section 3.3. This suffers from a significant deficiency. Few cities follow the standard urban model that generates a symmetrically monotonic rent gradient.[15] While there are several reasons for this, a significant one is accessibility. The rent gradient of the standard urban model is based on transportation costs, not distance. However, in actuality, accessibility of a location plays a significant role in its value and non-uniform variation in accessibility will generate a non-uniformity in the rent gradient.

The urban producer theory of chapter 2 relies upon space which does not vary in quality. However, in application, space varies significantly in its quality or productivity, as well as the additional services and amenities that may be included in the rent. To control for these additional elements of an office space, I turn to the literature on hedonic regressions. The basic idea is to estimate a hedonic model of the rents in the Cleveland CBD and then predict those rents for a specific type of property. This relies on the assumption that a competitive commercial real estate market will yield prices that are equal to the marginal product of the space to its tenants. Therefore the result is that I obtain a rental rate on a constant quality-productivity space. The variation then observed across the data reflects variation in the rent while holding the characteristics of the space constant.

While the hedonic regression can adjust for differences in the characteristics of the space, it cannot account for any asymmetry or complex nonlinearities in the rent gradient of the Cleveland CBSA. To address this issue I use least squares to estimate a locally weighted regression (LWR), also known as lowess or loess regression. The basic idea behind LWR is to allocate greater weight in a regression to observations that are closer to the point of reference. Traditionally this point of reference has

[15]See, for example, McMillen (1996) and Colwell and Munneke (2009).

been a point in space and the 'closeness' would be based on the euclidean distance from the point of interest. The method has also been applied to problems that are not geographic. In these instances, a multivariate set of characteristics define the location of the reference point and the observations in the dataset. A distance is then calculated based on these characteristics. The method has been used in the urban economics discipline by several authors including Meese and Wallace (1991), McMillen (1996), and McMillen and Redfearn (2010).

3.5.1 Details of the Locally Weighted Regression

A number of choices present themselves when formulating a locally weighted regression: 1) the choice of bandwidth, 2) the weighting function or kernel, 3) the dataset, 4) the functional form of the regression, and 5) the explanatory variables of the regression.

Bandwidth

The bandwidth defines which 'nearby' observations will be used in the estimation of the model parameters for a particular location. As discussed by Cleveland et al. (1988), bandwidth selection involves a trade-off of bias for variance. As the bandwidth increases, additional bias is introduced because observations derived from a different, nearby, data-generating process are used to estimate the local data generating process. However, as the bandwidth grows, additional data are used in the estimation leading to a reduction in the variance of the estimates. McMillen (1996) uses a bandwidth of 30%, while McMillen and Redfearn (2010) use 25%, meaning that the closest 25% or 30% of the data to the location of interest is used in estimating the expected value

of the dependent variable at that point.[16] In the analysis that follows, I maintain a bandwidth of 30%. Appendix C examines the variability in the analysis over various bandwidths. Most bandwidths support the urban producer hypothesis and those near to 30% provide predicted effects that are close in value.

Weighting

The weighting function converts the distance measure to a weight to be used in the individual regression calculations. The tri-cube function is a common weighting function. For an analysis of I data points with distances $\{d_1, d_2, \ldots, d_I\}$ from the location of interest, the tricube weight function takes the form $w_i = \left(1 - \left(\frac{d_i}{\max\{d_1,\ldots,d_I\}}\right)^3\right)^3$. I adopt this weighting function for all analysis. It has a number of desirable properties as pointed out by Cleveland et al. (1988). In addition, as mentioned by McMillen and Redfearn (2010), analysis tends to be relatively robust to the choice of the weight function because most weighting functions share the characteristic that they are decreasing in the distance, d_i. A final note is that observations at the maximal distance of the bandwidth receive a weight of zero and effectively shrink the bandwidth with the tri-cube function. When there are numerous observations at the same distance this can effectively provide a noticeable reduction in the sample size. This effect can be seen in table 3.16 where the number of observations in a regression varies from 790 to 898.

[16] Meese and Wallace (1991) is an exception to these ranges as they examined a range from 30% to 100%, ultimately using a bandwidth of 100%. Their choice was driven by the need to ensure that they were able to identify the parameters of interest in each of the individual regressions rather than the bias-variance tradeoff, though they report that bandwidth had little influence on their results. I explore the issue of bandwidth and parameter identification in the discussion of the form of the hedonic regression.

Data

The matched property method of section 3.4 relied on the use of property average asking rents for an attorney's location. In estimating the asking rent at a location it is not necessary to collapse the space data down to the building or property level. Instead, each space can be used as a separate observation. Use of the space data in the hedonic regression allows a greater amount of data without averaging out the variation in space characteristics. However it also has the effect of placing large weight on buildings with large numbers of spaces available for lease. The property with the most spaces for lease in the CoStar data is 26250-26300 Euclid Avenue in Euclid. This property has 168,000 square feet of rentable space and 67 leases available.[17] Comparing this to the largest property, in terms of rentable space, illustrates the problem. The Key Bank Tower at 127 Public Square, which I have defined as the center of the central business district, has 1.3 million square feet of rentable space. However only 9 leases are currently being marketed. I have chosen to use the rents of individual spaces rather than a property average. Different spaces within the same building will have different characteristics such as the services included in the rent and square footage of the space. It is important to control for this element in the analysis.

Explanatory Variables of the Hedonic Regression

The characteristics that are used to explain the rent must be selected in order to accurately predict the floor area rent at each attorney location. Several issues arise with respect to the specification.

[17] While this amount of available space would not be unusual for a property that was recently constructed or renovated, this property meets neither of those characteristics. While the CoStar data does not provide sufficient information to determine an exact reason, it is possible that this is the result of a large tenant vacating the property and the landlord choosing to offer to sub-divide the vacant space.

A note should be made about the form of the hedonic regression. My purpose here is to adjust for productivity. I do not want to also subtract out the land costs associated with a location as this will generate the most significant variation in asking rents. Therefore care must be taken in constructing the hedonic regression. Property characteristics that are strongly correlated with land values and weakly correlated with the productivity of the building need to be excluded. The coefficients on these variables will suffer from omitted variable bias due to the omission of land prices from the regression. I do not want to subtract out the land values. A prime example of a property characteristic that is a problem is building height. The tallest buildings are located in the Cleveland central business district where land values are commonly the highest. The average building height within one mile of the CBD is 10.9 stories, while the average outside this one-mile radius is 2.4 stories. This will lead to a strong correlation between land values and building height, though the height of a building is likely to have a minimal impact on the productivity of the businesses within. Distance from the CBD is another explanatory characteristic that is highly correlated with land values. I do not exclude distance from the CBD from the hedonic regression. Unlike building height, I do possess information on the distance from the CBD for each attorney location. Therefore when predicting the rent at a location I can add back in the effect of distance, including the indirect effect due to the omission of land rent, and thereby not remove the effect of land rent through the prediction process. Because I do not possess information on the building height at attorney locations, I cannot perform a similar adjustment for building height.

The bandwidth can also complicate estimation, as discussed by Meese and Wallace (1991) with respect to indicator variables. I find a similar problem with respect to energy efficiency and building age where it is the case that the subsample for particular regressions do not include any observations with the particular characteristic. In

this case the parameter of interest is not identified and its influence on productivity cannot be controlled for. I therefore do not include a control for "green" certified buildings. A similar problem arises with respect to controlling for building age using a set of indicator variables such as that in Eichholtz et al. (2010) . Building age is an important determinant of rent and therefore cannot be excluded. Indicator variables have traditionally been employed because the marginal effect of building age has been found to be decreasing in building age. While one might presume to use the natural log of building age to generate a similar decreasing marginal effect, that is generally ill-advised due to the presence of recently constructed buildings with an age of zero years. Adding a constant to building age prior to a logarithmic transformation would clearly generate biased estimates of the marginal effect of building age on the asking rent. However, the purpose of this exercise is not to estimate marginal effects. I am simply engaged in a forecasting exercise to accurately predict the floor area rent at a location. Therefore I add a constant equal to one to building age and rather than apply a series of indicator variables to control for age I use a continuous measure. This eliminates the issues with respect to identification that arises due to a non-uniform spatial distribution of building characteristics.

Finally, the form by which the distance from the CBD enters into the regression is important. If the rent gradient in Cleveland were a smoothly monotonic function, then a simple linear or logarithmic form for the mileage would suffice. However, as demonstrated by McMillen (1996), rent gradients are not necessarily monotonic and may have a number of sub-centers that exhibit local peaks in land prices. Therefore I add flexibility to the functional form by using a linear spline to model the effect of distance from the CBD. For each regression, a 9-knot (10 line segments) spline is used where the knots are placed so as each segment will have the same proportion of observations. The proportional distribution is necessary to avoid the issues of

identification, such as with building height, that would occur if the knots were set at fixed locations.

The independent variables of the hedonic regression, as well as the values used to predict the rent at each attorney location, are presented in table 3.15.

Functional Form

Hedonic regressions on residential housing prices commonly use a log-linear form in which the housing price (whether rental rate or sales price) is transformed using the natural log. Hedonic models of office rents have used both linear and log-linear functional forms. Eichholtz et al. (2010) adopt a log-linear form while Dunse and Jones (1998) report that they find a linear form rather than log-linear performs better. Bollinger et al. (1998) report that they find highly similar results between linear and log-linear forms of a hedonic model of office rents. I examine the performance of the log-linear and linear forms in order to choose between the two forms. This is done by estimating each rent regression using a Box-Cox transformation of the dependent variable. The Box-Cox method uses maximum likelihood to solve a nonlinear conditional mean function for the normal distribution where the dependent variable, y, is transformed as $y^* = \frac{y^\lambda - 1}{\lambda}$. λ is simply another parameter to be estimated. When $\lambda = 1$, then a linear form of the dependent variable is appropriate. When $\lambda = 0$, then a log transformation of the dependent variable provides a better fit of the data. And when $\lambda = -1$, then the reciprocal of the dependent variable is more appropriate. I find that for 63.0% of the lowess Box-Cox regressions I cannot reject the null hypothesis that $\lambda = 1$ at the 95% level of confidence. Conversely, I cannot reject the null hypothesis that $\lambda = 0$ for 16.9% of the regressions and I cannot reject the null hypothesis that $\lambda = -1$ for only 0.3% of the regressions. Consequently I adopt a linear form for the hedonic regressions.

3.5.2 Predicted Rents in the Cleveland CBSA.

Locally weighted hedonic least squares is used to estimate the productivity-adjusted rent at each location where an attorney is employed. The nearest 30% of the spaces offered for lease to an attorney's location are selected for analysis and the weights are constructed for those observations using the tri-cube function. The weighted hedonic regression is run on asking rents for individual office spaces using a set of space and property characteristics including a linear spline for the distance from the CBD. The resulting coefficients are used to predict the rent at a location with a fixed set of characteristics. The values used to predict the rent at a location are listed in the final column of table 3.15. This procedure is repeated for each attorney location in the dataset to obtain a productivity-adjusted rent at the attorney's location.

The office rent gradient in Cleveland can be visualized as a contour map. I use the locally weighted regression method to predict the rent at 201,347 points on a 0.1 x 0.1 mile grid within the boundaries of the Cleveland CBSA. Figure 3.12 displays the contour map of the predicted rents in the Cleveland CBSA, while figure 3.13 zooms in on the area within 5 miles of the CBD, where over half of the attorneys in my analysis are located. The location with the highest predicted rent is the center of the CBD and the rent for the fixed quality office space is predicted to be $24.38 per square foot per year.[18] The location with the lowest predicted rent, $4.78 per square foot per year, is on the southwestern edge of the CBSA near New London. The rent contour of Cleveland does exhibit the multi-peaked surface similar to that found in Chicago by McMillen (1996). For example, the peak about 12 miles east of the CBD is Beachwood, Ohio. In addition to being an affluent suburb of Cleveland it is

[18]While the asking rent in the CoStar space data have a range of $1.06 to $45.45, the 99th percentile is $24.50 per square foot per year.

home to several corporate headquarters.[19] To the southwest of the CBD exist peaks associated with Cleveland-Hopkins International Airport and the city of Strongsville, another affluent Cleveland suburb.

One anomaly is readily apparent in figure 3.12. Approximately 32 miles southeast of the CBD, estimated rents spike up to $18.42 per square foot per year. This illustrates one important point in regards to lowess estimation. There are no nearby properties in the CoStar data with which to estimate this location. The nearest office space for rent is 10 miles away. As such, the confidence one might have in this estimate is limited.

3.5.3 Testing Urban Producer Theory

Urban producer theory is tested using an ordered logit model with attorney quality as the categorical dependent variable and the predicted rent at the attorney's place of employment as an independent variable. In order to accurately test the hypothesis it is necessary to have consistent estimates of the covariance matrices of the rent estimation regressions and of the ordered logit model. One area of concern in regards to the consistent estimation of the covariance matrix is the assumption that the error terms are independent and identically distributed (i.i.d.). For example, the ordered logit model uses information on more than 8,000 attorneys, however these attorneys are located at fewer than 2,000 locations. It is certainly possible that the error terms for attorneys employed at the same location are not independent. The covariance matrix can be adjusted for this clustering by allowing the error terms of observations at the same location to be correlated.[20] All hypothesis tests reported for the ordered logit models are based on cluster-corrected standard errors.

[19] See `http://www.beachwoodohio.com/econdev.html` for a list of significant corporations located in Beachwood.

[20] See Angrist and Pischke (2009, chap. 8) for a discussion of clustered standard errors.

A similar issue arises in the rent regressions. There are observations on multiple spaces for lease in the same property. Many of the characteristics of the space, such as vacancy rates, building size, building class, and location, are the same for all spaces in the property. This can lead to correlation among the error terms of spaces located in the same property and commonly would be accounted for by using cluster-adjusted standard errors. Unfortunately, an additional complication enters the mix for the rent regressions. The rent regressions use locally weighted least squares as proposed by Cleveland et al. (1988). This method assumes that the error terms are i.i.d. There has not been any asymptotic theory developed that would suggest that appropriate way in which to treat the weighting process in the development of a consistent estimate of the covariance matrix in the presence of non-i.i.d. errors. Therefore all of the results from the rent estimation are based upon standard errors that are consistent under the i.i.d. assumption. Appendix B examines the sensitivity of the tests of urban producer to the assumptions regarding the error terms of the rent regressions. The results indicate that the assumption regarding i.i.d. errors in the rent regression do not significantly affect the hypothesis tests.

The final issue to address is the effect of using the predicted rent at an attorney's location rather than the actual rent at the location. It is a prediction and therefore is measured with a known error distribution and the standard method for calculating the covariance matrix of the estimation would be biased because it does not account for the fact that the independent variable was measured with error. Murphy and Topel (1985) develop a method to correct variances for this two-step estimation procedure, while Hole (2006) provides guidance in implementing the correction in modern statistical software. The adjusted covariance matrix of the second step ordered logit

model is given by

$$V_2 + V_2 \left(C V_1 C' - R V_1 C' - C V_1 R' \right) V_2 \tag{3.4}$$

where,

$$V_1 = E\left[\frac{\partial L_1}{\partial \theta}\frac{\partial L_1}{\partial \theta'}\right]$$
$$V_2 = E\left[\frac{\partial L_2}{\partial (\alpha, \beta)}\frac{\partial L_2}{\partial (\alpha, \beta)'}\right]$$
$$C = E\left[\frac{\partial L_2}{\partial (\alpha, \beta)}\frac{\partial L_2}{\partial \theta'}\right]$$
$$R = E\left[\frac{\partial L_2}{\partial (\alpha, \beta)}\frac{\partial L_1}{\partial \theta'}\right]$$

θ are the estimated coefficients of the first stage model

(α, β) are the estimated coefficients of the second stage model

The sample analogs of the expectations can be used to calculate the estimates. L_1 and L_2 are the likelihood functions for the first and second steps and therefore V_1 and V_2 are the estimated covariance matrices for the first and second steps of the estimation strategy.[21] I assume independence of the first step and the second step of the estimation which implies $R = 0$.[22] The data for each stage are from entirely different sources and correlation is unlikely to exist. Finally, I must account for the locally weighted estimation method in the first step. Murphy and Topel (1985) discuss how to adjust the standard errors when different samples for different groups of the second step observations are used to estimate the first step parameters. Essentially

[21] The first step of the estimation procedure uses least squares, however this is equivalent to maximum likelihood estimation under a normality assumption, see Amemiya (1985, sec. 1.3), for example.

[22] In the case of independence between the first and second steps it is the case that the uncorrected standard errors are biased downward.

the term CV_1C' becomes a weighted average of the separate first step regressions. Under this procedure the adjusted variance becomes

$$V_2 + V_2 \left(\sum_{s=1}^{S} \frac{n_s}{N} C_s V_{1s} C_s' \right) V_2 \tag{3.5}$$

where s indexes the separate first step regressions for each of the 1,981 attorney locations, N is the total number of attorneys, n_s is the number of attorneys at location s, and

$$V_{1s} = E\left[\frac{\partial L_{1s}}{\partial \theta_s}\frac{\partial L_{1s}}{\partial \theta_s'}\right]$$
$$V_2 = E\left[\frac{\partial L_2}{\partial (\alpha, \beta)}\frac{\partial L_2}{\partial (\alpha, \beta)'}\right]$$
$$C_s = E\left[\frac{\partial L_2}{\partial (\alpha, \beta)}\frac{\partial L_2}{\partial \theta_s'}\right]$$

3.5.4 Results of the Estimation

First Step Results

The first step of the estimation procedure consists of lowess regressions on 1,981 locations where an attorney in the data was employed. Table 3.16 provides a number of summary statistics of the lowess regressions. Productivity adjusted predicted rental rates range from \$5.76 to \$24.38 per square foot per year for a 1,975 square foot space in a 37 year old class B building.[23] This can be contrasted with the unadjusted range of \$1.06 to \$45.45 for space rents in the CoStar dataset. Examination of the predicted rents indicates that the maximum predicted rent occurs at the center of the CBD, as one might expect, though, as illustrated by figure 3.12, there are high rent areas

[23] See table 3.15 for a complete description of the base characteristics used to obtain the predicted rent.

located well outside the CBD. This fact, in conjunction with the range of rents at attorneys' locations assists in identifying the input quality gradient.

The fit of the models, as measured by the unadjusted R^2, ranges from 0.35 to 0.78, with a median value of 0.52. This compares to values ranging from 0.63 to 0.79 found by Bollinger et al. (1998), Dunse and Jones (1998), and Eichholtz et al. (2010). My failure to include building height in the rent regressions may explain the slightly poorer fit when compared to these works. I also categorize the outcome of the bandwidth selection process. Each lowess regression uses 30% of the CoStar office space data nearest to the attorney's place of employment. The distribution of the minimum and maximum distance from the attorney location to the office space locations is presented. Fully half of the lowess regressions have at least one space with asking rent data located within one-third of a mile of the attorney's location and more than three-quarters have at least one rental data point located within a mile. However, there are a few attorney locations that are not near any of the CoStar rental locations. There is a single attorney that is over 10 miles away from the nearest data point on office rents. While these nearby rental data point have the largest weights in the regressions, more distant observations still contribute to the estimation. The most distant datapoint used is over 38 miles from the attorney's location. However for half of the regressions the most distant rental datapoint is no more than 10 miles away. In addition, because an attorney location may hold multiple attorneys — particularly locations near the CBD — the extremes make up a small portion of the attorney sample. Fully 55% of the attorneys in the estimation have a predicted rent based on a lowess regression in which all office spaces used to estimate the rent at their location are within 5 miles of the attorney's location. Only 6% of the attorneys have a predicted rent based on a lowess regression where the nearest office space used in the estimation is more than one mile away. Therefore,

the problems associated with predicting rent using distant data, such as was found in the far southeast of the Cleveland CBSA are of a limited concern.

Each regression generates a set of estimated coefficients for each of the characteristics contributing to the space rent. I conduct a standard hypothesis test for each coefficient in each regression. Three outcomes for each estimated coefficient are possible: 1) failure to reject the null hypothesis that the estimated coefficient is equal to zero, 2) rejection of the null hypothesis that the estimated coefficient is equal to zero and the point estimate is positive, and 3) rejection of the null hypothesis that the estimated coefficient is equal to zero and the point estimate is negative. Table 3.17 presents the results of this hypothesis testing of the coefficients for each of the 1,981 regressions. While more formal tests of multiple hypotheses are available, just simple observation will suffice for my purposes. The fraction of the regressions falling into each of these three categories is indicated. I commonly observe higher rents for class A and B buildings than the default class C buildings. Older buildings tend to have lower rents. Not surprisingly, offices offered for lease on a net basis or with a limited amount of services included in the lease rent for lower amounts than offices that include all utilities in the asking rent. These results can be compared to the global regression of table 3.11 as well as the results of Bollinger et al. (1998), Dunse and Jones (1998), Eichholtz et al. (2010), and Fuerst and McAllister (2011). The signs of the median coefficients for the lowess regressions do match with those of table 3.11. However a few curious instances still exist. For example, the global regressions finds statistically significant effects for building that have been renovated and have amenities. However, from examining table 3.17, there appears to be little evidence of this effect. The effect of amenities on the rent of commercial office spaces varies with Eichholtz et al. (2010) finding a statistically significant effect and Bollinger et al. (1998) finding mixed results for the various types of amenities. Eichholtz et al. (2010)

do not find that renovation adds to the rent of commercial buildings, though Dunse and Jones (1998), when interacting this with the building age do find a positive effect. Previous findings of the effect of the size of the office space on the rent are mixed as well. Dunse and Jones (1998) find that the size of the space for lease has a positive effect on the asking rent for the office space, while Fuerst and McAllister (2011) find a negative effect. I find very little evidence of a positive effect, however the evidence for a negative effect is only slightly stronger.[24] Overall, previous hedonic estimates of the determinants of commercial rents have found mixed results with a handful of characteristics such as lease type, building class, and building age providing consistent results across studies. My hedonic regressions mirror these standard conclusions.

Second Step Results

Table 3.18 presents the results of the ordered logit estimation of the effect of fixed-quality office rents on the quality of the attorneys employed at a location. The results of model 1 indicate that office rents have a positive and statistically significant effect of the quality of attorneys at a location. This result provides strong support for the primary hypothesis generated by urban producer theory.

While the estimated coefficient does indicate that rent has a positive effect on attorney quality, it does not provide an indication of the magnitude of the effect. Because I only observe discrete quality categories rather than the underlying cardinal quality measure, no relevant scale to measure the magnitude is conveyed in the parameter estimate. Therefore, as with the previous estimation strategies, I examine the effect of rent on the probabilities of observing high quality and low quality attorneys at a location. Figure 3.14 illustrates the effect of office rent on the probabilities and elasticities of low and high quality attorneys, as estimated by model 1. At the CBD,

[24]Neither Bollinger et al. (1998) nor Eichholtz et al. (2010) include the size of the space for lease in their analyses because their basic unit of observation is the building.

with an estimated rent of $24.38 per square foot per year, I am 4.2 times more likely to observe a high quality attorney than I am to observe one at a location with a rent of half that amount. Conversely, I am 4.1 times less likely to observe a low quality attorney at the rent levels found in the CBD than at a location with a rent of $12.19 per square foot per year. Regardless of the office rent, low quality attorneys are more commonly observed than the high quality, Yale-educated, attorneys, though in the buildings with the highest rent, the probabilities of observing high and low quality attorneys are much closer than at lower rent levels. Estimated elasticities range from 0.6 to 2.9 in absolute value. At the higher rent levels, a 1% increase in space rents leads to a 2.9% increase in the probability of observing a high quality attorney at that location.

As with the evaluation of the estimation that uses matched properties, I examine several alternative specifications of the ordered logit model. Table 3.18 presents these results. Model 2 introduces the distance from the CBD into the model. The estimated effect of rent on attorney quality remains positive and statistically different from zero. As with the matched model, the effect of distance fom the CBD on attorney quality cannot be statistically distinguished from zero. Figure 3.15 illustrates the estimated elasticities and probabilities of low and high quality attorneys at the median and mean distances from the CBD. These values are of a similar magnitude to those of model 1.

Models 3 and 4 introduce the size of the attorney's employer into the estimation. This has a noticeable impact on the estimated coefficients, as reported in table 3.18. As illustrated by figures 3.16 through 3.18, the elasticity of the attorney probabilities are now uniformly less than 2 in absolute value. For example in models 3 and 4, the elasticities of the attorneys range from 1.26 to 1.34 in absolute value at a rent of $24.38 per square foot per year, however this range was 2.49 to 2.89 in models 1 and

2. It is clear that the size of a firm has a substantial impact on the quality of the attorneys that are employed.[25]

Figure 3.19 compares the effect of rent on the elasticities and probabilities of low and high quality attorneys at the median values of distance from the CBD and firm size. These comparisons highlight the limited role that distance from the CBD plays in attorney quality. In the figure models 3 and 4 are virtually indistinguishable, while models 1 and 2 are also reasonable similar.

Evaluating Law School Ranking as a Measure of Worker Quality

Law school ranking is admittedly a rough estimate of attorney quality. The lowest quality graduate from the highest ranked law school (Yale) may well be of a lower overall quality than the highest quality graduate from the 20th ranked law school (Minnesota). These differences may not be readily apparent to firms when they are hiring graduates with a limited amount of work experience. If this is the case, then I should observe weaker sorting among more experienced attorneys because employers are able to observe a more accurate representation of the attorneys' quality independent of their educational achievement. The sorting will be stronger with less experienced attorneys because the educational achievement is an important piece in a firm's evaluation of the quality of an attorney. In addition, to the extent that experience contributes to the human capital of an attorney, the relationship between attorney quality and law school ranking may be weakened further. It may well be the case that an attorney with 10 years of experience that was trained at the 50th ranked law school has the same level of human capital as an attorney with no experience from

[25]While a theory explaining why large firms hire high quality attorneys is beyond the scope of this dissertation, I can offer a few observations. One possibility is that larger firms are more productive (possibly measured as success in litigation) than smaller firms and they signal this higher productivity by locating in high rent properties and hiring attorneys from prestigious law schools. Another possibility is that the production process in large firms is such that they are able to obtain efficiencies from high quality attorneys that cannot be captured by smaller firms.

the 20th ranked law school. I test this proposition by examining whether there are differences in the attorney quality gradient based on the experience of the attorney. I separate the attorney sample into attorneys that are experienced and attorneys that are inexperienced. I define an experienced attorney as one that has been a member of the Ohio Bar for at least 6 years, while an inexperienced attorney has been a member of the bar for less than 6 years. In addition, I create a further subsample consisting of highly experienced attorneys that have been members of the Ohio Bar for at least 20 years.[26] Table 3.19 presents the results of the ordered logit model on the subsamples, as well as the base model 4. Figure 3.20 clearly demonstrates that the attorney quality gradient is steeper for more experienced attorneys. In fact, for inexperienced attorneys, there is no statistical support for the existence of an attorney quality gradient. This does not support my working hypothesis that employers have a cleaner signal on the productivity of more experienced workers than of less experienced workers. However, additional factors may be at work. Garicano and Hubbard (2005) find that cognitive ability, as measured by law school quality, and experience are not substitutes in production. Additionally, inexperienced attorneys, commonly referred to as associates, are likely to perform the same routine legal tasks under the supervision of an experienced attorney with little need for advanced reasoning skills. In this situation, graduates from lower quality law schools may be reasonable substitutes for graduates from higher quality schools. Once attorneys become more experienced and are trusted with complex legal tasks, their inherent abilities will differentiate the low quality from the high quality practitioners. Therefore, the effect I observe here may be more closely related to the higher productivity associated with

[26]Ideally I would like to identify inexperienced attorneys as those who have not yet reached the point where they are considered to become partners of a law firm. However, as Raasch (2007) indicates, the partnership track varies by firm and ranges from 6 to as much as 12 years. Therefore I have chosen the first cutoff to represent the lower range of the partnership track and include the second cutoff to further evaluate the effect of experience.

more experienced attorneys and the ease of substituting low quality for high quality inexperienced attorneys rather than a rejection of hypotheses regarding uncertainty over worker quality.

A similar argument in regards to uncertainty over worker quality may hold with respect to the location of the law school attended by the attorney. Firms located in Cleveland are likely to have stronger networking ties to local law schools and therefore acquire additional signals regarding the quality of potential employees. Under this hypothesis I would expect to find weaker sorting among attorneys from Ohio law schools than from out-of-state law schools. Table 3.20 presents the results on the ordered logit estimation on subsamples of in-state and out-of-state graduates, as well as the full sample of Cleveland attorneys. The estimates provide support for this hypothesis. Interpreting the parameters of the ordered logit is problematic in this instance. While the results analyzing the effect of attorney experience on the attorney quality gradient were based on the full range of attorney quality levels, that is not the case here. The law schools, and therefore attorney quality levels, are not the same in the two subsamples. Therefore, calculating the probabilities of high quality attorneys to evaluate the attorney quality gradient makes little sense as the high quality attorneys are of different qualities in the two subsamples. In the out-of-state subsample, the high quality attorney is from Yale, the top ranked law school in the nation, while the highest quality in-state attorney is from Ohio State University, a school ranked 35th in the nation. One natural comparison may be to the probabilities of an in-state school to that of a closely ranked out-of-state school, such as a comparison between the probabilities of observing an Ohio State graduate and an University of Wisconsin graduate, since both schools are ranked 35th in the nation. This is ill-advised. Because of the nature of the ordered logit and the role of the ancillary parameters, it is the case that the University of Wisconsin may have

a probability that is either decreasing or increasing in the rent for the out-of-state regression, while in the in-state regression, Ohio State is guaranteed to be increasing in the rent when the coefficient on rent is positive because it is the highest ranked in-state school.

A similar, though possibly less severe, issue exists with respect to evaluating the probability of observing a low quality attorney. Unlike the highest ranked law school, there are many law schools in the lowest ranked tier. If all of these schools are considered to be equally poor by employers, it may be possible to test the proposed hypothesis. However, if employers differentiate between the schools, and it is only US News that considers them to be equal, then little is solved. While comparing the predicted probabilities of the subsamples is inappropriate because the ranges of possible outcomes are different, comparing the elasticities may be less so. At a minimum, the elasticity will account for base probabilities and therefore minimize any distortion caused by the different number of quality levels.

I proceed based on the assumption that employers themselves cannot distinguish between the lowest ranked schools. The results are supportive of the hypothesis that employers gain better information about attorneys that attended schools within the state of Ohio. As shown by figure 3.21, the attorney quality gradient is more elastic with respect to out-of-state low quality attorneys than it is with respect to in-state low quality attorneys, possibly due to the additional information Cleveland employers can easily access with regards to the complete quality of an in-state graduate.

Formal statistical testing of these hypotheses has not been performed. The traditional method of testing hypotheses such as these would be to nest the models or include interaction terms so that estimation is performed on the full sample. In my application, this leads to a problem in obtaining consistent estimates of the covariance matrix. For example, in order to test the hypothesis that experience attenuates the

attorney quality gradient, I would include an additional regressor consisting of the interaction between experience and rent. Because rent is an estimated value from the first step regressions, the covariance matrix must be adjusted for this. However there is no clear guidance on the adjustment that must be made for the case of an interaction term. Fortunately, in the case at hand, the conclusions of the two extensions are quite clear and provide confidence in the robustness of the statistical conclusions. Inexperienced attorneys do not have a statistically significant quality gradient while experienced attorneys do. Similarly, in-state graduates do not have a statistically significant quality gradient while out-of-state graduates do. However, this does not equate to a formal statistical test of the difference between the estimated coefficients on rent in the regressions.

3.6 Summary

This section has implemented a series of empirical tests of urban producer theory as developed in chapter 2. The basis of these tests is the prediction of urban producer theory that an input quality gradient will exist in response to the well-documented land rent gradient. Using information on rents and attorneys in the Cleveland CBSA I am able to demonstrate that the probability of a location having a high quality attorney employed there is increasing in the floor area rent. These findings are robust to the manner in which rent is calculated. In addition, the results are robust to the inclusion of other factors which may also contribute to the sorting of high quality attorneys into high rent areas such as central business districts. Estimates of the elasticity of the probability of a high quality attorney vary from 0.77 to 1.71, depending upon the specification, at the median floor area rent of attorney locations in Cleveland of $14.41, with higher values at higher floor area rents.

Evidence that higher quality attorneys sort into central business districts is found by Limehouse and McCormick (2011). They find a significant correlation between law firm quality, as measured by total revenue per lawyer, and the decision to locate in the CBD. They report that average revenue per lawyer is about 13% higher for firms located in the CBD. Their unstated assumption is that revenue per attorney is a measure of the quality of the input rather than reflecting a locational premium in the price of the output. Because rents in central business districts tend to be higher than in outlying areas, this result is consistent with urban producer theory. My analysis stresses this hypothesis further by accounting for the existence of high rent subcenters such as Beachwood, Ohio. My findings indicate that while locating in the CBD plays some role in firm quality, it is really the floor area rent that explains the sorting of high quality attorneys.

My results provide a contrast to the findings of Fu and Ross (2013). They find that workers within a metropolitan area do not sort into areas with high employment concentrations based upon unobserved differences in ability. Fu and Ross use confidential data from the long form of the 2000 Decennial Census to examine the relationship between wages, as a measure of worker ability or productivity, and employment density. Their main focus is on determining whether the wage premia observed among workers in dense employment locations is a locational premium due to the density or a premium due to the ability of the workers. The difficulty with an exercise such as this is that it is difficult to observe fully the ability of a worker. In this dissertation I use a more refined measure of ability by relying on the quality, rather than quantity, of education of the worker. Fu and Ross rely on residential location that theory predicts will be based on sorting and has been empirically verified to identify the the productivity of workers. This residential sorting yields a situation whereby workers residing near one another have similar expected lifetime earnings, net of commuting

costs. Therefore the inclusion of residential fixed effects will sweep out unobserved productivity differences and the remaining correlation between wages and employment density will be due to differences in commuting costs and premia associated with the density of the work location. They find that the inclusion of residential fixed effects have minimal effects on the estimates of the wage premium leading them to conclude that sorting into high density work employment centers based on unobserved ability is not a significant source of bias.

To the extent that employment density and rent are correlated this contrast between my work and that of Fu and Ross (2013) leads to conflicting conclusions. Several significant differences should be pointed out. Identification of the sorting effect is of a completely different nature. In my work, I obtain additional information on the ability of a worker that would normally be classified as unobservable. Fu and Ross on the other hand rely on an assumption that residents of census tracts have the same ability. While I use a very specific occupation heavily concentrated in a single industry in a single metropolitan area, they use a much broader sample. Fu and Ross use a sample of workers from a broad range of occupations, industries, and metropolitan areas and use fixed effects so that estimation is based upon the within-group average across the groups. If it is the case that the urban producer effect varies substantially between industries, such as might be generated by different parameters in the production function, the ability to measure such an effect is weakened. The greater the employment in industries with weak sorting, then the less the estimated effect will be. In a similar fashion, if the relationship between rent and density varies substantially between metropolitan areas then the relationship hypothesized by urban producer theory would be unrelated to density. Finally, Fu and Ross measure density at the level of the Public Use Microdata Area (PUMA). If it is the case that a wide range of rent levels reside within the PUMA, then again the relationship between density and

the sorting mechanism of urban producer theory is broken. There were 18 PUMAs in the Cleveland CBSA in the 2000 Census. However 12 of these areas were located in Cuyahoga county where most of the attorneys in my sample are located, though I am unable to determine the extent to which these areas exhibited variations in rent.

In the chapter that follows, I examine the implications of the demand-side sorting based on input quality that is implied by urban producer theory. Several avenues of research in urban and regional economics are examined and the potential biases that urban producer theory might introduce are discussed.

Chapter 4

Implications of Urban Producer Theory

This chapter examines some implications of urban producer theory for current research in two areas. First, there is a literature that measures intra-urban wage gradients which are supposed to reflect wages paid to identical workers at various employment locations throughout the city. Second, I develop implications for the vast literature attempting to measure and then attribute agglomeration economies to characteristics of areas within cities.

4.1 The Intra-Urban Wage Gradient

The intra-urban wage gradient refers to the relation between the wages paid by employers and the location of employment within a city. The standard monocentric urban model of Muth (1969) and Mills (1972) generates a negative relationship between the wage and the distance to the central employment center due to the need to compensate workers for more costly commutes associated with employment in the

CBD. In equilibrium, employment locations closer to suburban residential locations are able to attract workers with lower wage rates due to the reduced commuting costs associated with those nearby employment locations. As discussed by Timothy and Wheaton (2001), a similar, though non-monotonic, relationship will hold in a urban area with multiple employment centers.

A number of studies have verified the existence of an intra-urban wage gradient in particular cities, including Cleveland. Among them are Eberts (1981), Madden (1985), Hendricks (1989), McMillen and Singell Jr. (1992), and Ihlanfeldt (1992). The basic strategy of comparing the wages of workers in the central city to those outside the central city is a common thread in this work. The details of each estimate however varies based on the data available to each investigator. For example, Eberts (1981) and Hendricks (1989) do not observe the wages of individual workers but instead observe the average wage paid by particular employers (municipalities and hospitals, respectively). Others, such as Madden (1985), McMillen and Singell Jr. (1992), and Ihlanfeldt (1992) are able to observe the wages earned by the individual worker. Variation also exists in the ability to precisely identify the location of the workplace. Eberts (1981) and Ihlanfeldt (1992) for example are able to obtain reasonably precise estimates of the distance of the employment location from the CBD. In contrast, McMillen and Singell Jr. (1992) and Hendricks (1989) are limited to identifying workplaces as being in a central city or suburban location and therefore cannot estimate a wage gradient that is a function of a cardinal measure of distance. Ihlanfeldt (1992) treads a middle ground on these issues. He is able to obtain the general location of the place of employment relative to the CBD and at the same time has a representative sample of U.S. workers. Consequently he can estimate how wages fall as the distance from the CBD increases while controlling for a selection of worker characteristics.

In addition to the intra-urban wage gradient, there exists the rent gradient within urban areas. As demonstrated by Coulson (1991), McMillen (1996), and Haughwout et al. (2008), land values, and consequently rents tend to decline with the distance from the CBD. Urban producer theory posits that the rent gradient, in the presence of congestion in production, leads to an input quality gradient in urban areas. Because higher quality workers earn higher wages, this effect increases the size of the intra-urban wage gradient with an additional component based on the quality of the workers employed at various distances from the CBD. Therefore observed wage gradients consist of two components: 1) compensation for commuting, and 2) payments for higher productivity. Because the wage gradient traditionally has been considered solely as compensation for commuting it has been used as a proxy to examine commuting costs in urban areas. For example, Leigh (1986) reports that the compensation for commuting was \$2.54 per hour, or 38% of the average hourly wage based upon an estimate of the wage gradient.

The estimates of the input quality gradient for lawyers in Cleveland from this dissertation can be used to investigate the nature and size of bias introduced into estimates of the intra-urban wage gradient when details on input quality are ignored. I begin by generating a naive estimate of the wage gradient in Cleveland.[1] Ihlanfeldt (1992) provides a good example of the measurement of the intra-urban wage gradient based upon distance from the CBD and data from individual workers. I follow his method to the extent that it is feasible with my data. Ihlanfeldt (1992) explains the log wage as a function of worker characteristics and the log of the distance from the

[1]As demonstrated by figure 3.12 Cleveland does not exhibit a monotonic rent gradient as one might expect in a monocentric city. Accordingly, I also restricted the following analysis to the region within 5 miles of the CBD, a region in which rents are falling with distance from the CBD. In addition, as illustrated by table 3.4, over half of the private sector attorneys in Cleveland are located in this region. Restricting the sample to the region within 5 miles of the CBD does not change the qualitative findings that are reported below.

CBD.[2] Ihlanfeldt's individual characteristics include age, schooling, disability, past unemployment, veteran status, marital status, and sex. Only a handful of these characteristics, namely age and sex, are available in my data on Ohio attorneys. In addition, I have the years of experience of the worker, which is commonly proxied for by the difference between age and schooling. Of course I also have a measure of the quality of the attorney's training, which is information rarely available to researchers. Ihlanfeldt acknowledges the concern that may arise over the lack of information on worker characteristics that influence the wage rate and also lead workers to select themselves into employment closer to the CBD. He finds very little evidence of this with a test for selection. My data do not contain the wages earned by the attorneys. However, I do know the school attended and US News reports the 25th, 50th, and 75th percentiles of the private sector starting wages for graduates of each law school. Consequently, while Ihlanfeldt (1992) has information on the actual wages earned by a worker, I must use the median (or other quartile) wage earned by a worker in that profession with a degree from that law school. Unlike Ihlanfeldt (1992) I do not include attorney characteristics in my estimation of the intra-urban wage gradient because the wages are not specific to an attorney and therefore are not correlated with the characteristics of a particular attorney such as experience and sex. What these wages do represent is the spatial wage distribution that would exist for the median (or other quantile) graduate of the law school. This does present an advantage that individual characteristics that are uncorrelated with individual law schools, such as experience, cannot influence the regression. Of course, the quality of the school's graduates is not controlled for. This is exactly the point behind estimating the naive

[2]Because some workers are employed at the CBD, Ihlanfeldt adds 1 to distances in order to take logs. I follow his technique by adding this arbitrary constant in order to facilitate the transformation. This transformation complicates the estimated coefficients as they can no longer be interpreted as elasticities. See Johnson and Rausser (1971) for a discussion of the biases that arise from this adjustment.

wage gradient.

One problem with using the reported median income of graduates is that the data appear to be censored at $160,000. There is some debate over whether these values are over-reported by law schools as recently evidenced by Badenhausen (2011). Cloud and Shepherd (2012) provide a discussion of the ease by which law schools can manipulate the rankings, in particular with respect to employment information on their graduates. These fall into an ongoing debate of the relevance, and veracity, of law school rankings. Caron and Gely (2006) introduce an issue of the Indiana Law Journal devoted to a symposium on the controversy. Therefore there is some concern that the reported incomes do not reflect the actual income distribution. What remains unanswered is if the bias is simply a constant or multiplicative effect for all law schools or whether the bias is systematically related to the ranking or location of the school. An additional concern with using the median income for all graduates of the law school arises due to regional variations in wages. Regional wages are influenced by a number of issues, including the cost-of-living and the presence of amenities. If a law school's graduates are not uniformly distributed across the nation then the median wages reported for all graduates may not reflect the median wage that the graduates would earn in Cleveland. Table 3.3 demonstrates that local law schools supply a much larger fraction of the pool of lawyers than do more distant law schools. For example, Cleveland State has 3,363 graduates in the private sector of the Cleveland CBSA. According to US News, 90% of Cleveland State's graduating law class of 2008 were employed in the states of Illinois, Indiana, Michigan, Ohio, and Wisconsin, while only 7.3% of Harvard's graduating law class were employed in those same states. Clearly there is a regional effect at work here. Furthermore, a simple log-linear regression with the attorney data indicates that a 1% increase in the distance from the attorney's law school to the attorney's employment location

yields a 3% increase in the median starting salary. This should not be too surprising. Table 3.2 indicates that many of the highest ranked law schools are located at a significant distance from the state of Ohio. No Ohio law schools are among the 18 schools indicating that the median private sector starting salary of their graduates is $160,000. Therefore I expect a certain amount of bias to enter into this calculation as the wages of the highest quality attorneys are likely to be overstated relative to what they actually earn when employed in Cleveland.

Many of the studies estimating wage gradients have relied upon data that does not precisely identify the employment location of the worker. Commonly the employment location may simply be classified as suburban or urban within the metropolitan area. This eliminates the possibility of obtaining estimates of the wage gradient based upon cardinal measures of distance, though they are able to show that wages are lower in suburban areas than in the central city of the metropolitan area. Ihlanfeldt (1992) provides an exception to this situation thereby facilitating a comparison of my estimate of the naive wage gradient in Cleveland with the estimates contained in of Ihlanfeldt (1992, table 2). Ihlanfeldt reports the results of a log-linear regression in which a worker's wage is explained as a function of the worker's characteristics and the distance of the workplace from the CBD. My results are presented in table 4.1 and the least squares log-linear specification is strikingly similar to Ihlanfeldt (1992).[3] Ihlanfeldt's coefficients are -0.050, -0.041, and -0.042 for white professionals in Philadelphia, Detroit, and Boston, respectively. My estimated coefficient based on the median wage is -0.040 for attorneys of all races in Cleveland.[4] Ihlanfeldt reports

[3]Because wages and distance are in log form, differences in scale will be incorporated into the constant of the log-linear regression rather than the coefficients on independent variables.

[4]This estimated coefficient was generated by least squares to facilitate the comparison with Ihlanfeldt (1992). All other calculations from non-linear models in this section are based on a poisson model which simplifies prediction of the wage rate when a log-linear or semi-log functional form is assumed.

a reduction of earnings of between 4.8% and 18.9% when workers are employed 10 miles from the CBD. As demonstrated in table 4.2, I estimate that wages are reduced by 5.9% to 9.9% at this distance.

The fixed-rent wage gradient is obtained from the estimation described in chapter 3. In particular I use the results from models 2 and 4 as presented in table 3.18 to predict the probabilities of attorneys of each of the quality levels locating at a range of distances from the CBD while holding rent fixed at the median value in the data of $17.66 per square foot per year and firm size fixed at the median of 5 attorneys. The resulting probabilities are then used as weights in conjunction with the median starting private sector wage that is reported to be earned by graduates of the particular law school to calculate an average starting private sector wage at each distance form the CBD. Figure 4.1 illustrates the naive and fixed-rent wage gradients for the Cleveland CBSA while table 4.2 presents the estimated wages. The results clearly illustrate the bias that can be introduced by the sorting generated by urban producer theory. The naive estimate of the urban wage gradient suggests that the average wage 40 miles from the CBD is 14.9% lower than the average wage at the CBD. However, after accounting for the input quality gradient as estimated in model 2, the wage only falls by 4.2% over that distance. These results suggest that, at least among Cleveland attorneys, the input quality gradient accounts for more than two-thirds of the fall in observed wages over this distance. The disparity between the naive wage gradient and the fixed-rent wage gradient is greatest near the CBD and becomes less significant as the distance from the CBD grows.[5] The fixed-rent gradient from model 2 falls at nearly the same rate as the naive wage gradient once

[5]Some of this disparity may be due to the functional form selected for the naive wage gradient. The disparity for linear and semi-log forms of the naive wage gradient are greater than those reported for the log-linear form at distances in excess of 22 miles and smaller at lesser distances. Regardless of the functional form of the naive wage gradient, the fraction of the wage gradient due to the input quality gradient exceeds two-thirds.

the edge of the CBSA is reached. Near the CBD, where rental rates are high, there is a significant amount of sorting based on input quality and therefore a significant deviation between the naive wage gradient and the fixed-rent wage gradient. As the distance from the CBD increases, the rental rates fall and the amount of sorting also falls. As explained in section 2.1.3, the urban producer model with separable input quality is able to constrain the sign of the divergence based on the degree of substitution between inputs. These empirical results are consistent with a situation with relatively more elastic substitution between inputs.

4.2 The Measurement of Agglomeration Economies

Urban economics has been interested in the causes of agglomeration for decades. A plethora of explanations have been offered to explain why we observe economic activities concentrated in discrete regions rather than spread out uniformly across space. Among them are theories of natural locational advantages such as being located close to sources of natural resources or natural transportation corridors. While there is certainly evidence such as Davis and Weinstein (2008) that some portion of agglomeration is due to this effect, recently research has turned towards agglomeration economies. Agglomeration economies are an endogenous mechanism by which concentrations of economic activities create both positive and negative externalities in regards to the productivity of the inputs to production. Congestion, commonly resulting from density, is a leading cause of agglomeration diseconomies. A number of mechanisms related to density have been proposed for positive agglomeration economies. Among them are information sharing, improved matching of the inputs to production, and knowledge acquisition and spillovers. As discussed by Combes et al.

(2011), empirical investigations of the contributions of the various proposed mechanisms to productivity have largely focused on labor productivity, as evidenced by wages. The empirical strategy commonly has consisted of comparisons of the productivity across different geographic units such as metropolitan areas, counties, states, or other geographical or administrative collections. Identification of the size of the agglomeration effect is then accomplished based on differences in the observational characteristics of the regions. Combes et al. (2011) review a number of sources of potential bias in the measurement of agglomeration economies. Of particular interest to this dissertation is concern over the endogeneity of labor quality. They review several mechanisms whereby labor quality may be correlated with density including what they refer to as ability sorting. In response to concerns in regards to ability sorting among regions leading to biases in estimates of agglomeration economies, a number of studies have examined whether high-skill workers are sorted into dense urban areas. Combes et al. (2008) estimate the determinants of wages across French regions and find strong evidence that high-skill workers sort into areas of dense employment. They find that up to half of the correlation between wages and density may be due to the sorting of high skill workers into urban areas. Mion and Naticchioni (2009) examine the spatial distribution of wages in Italy. They find that nearly 75% of the variation in wages across city sizes is due to sorting of workers' skills. Within the U.S. numerous studies provide evidence that high skill workers sort themselves into urban areas. The most straightforward evidence is simply the observation that the ratio of college educated to non-college educated workers is increasing in city size (see for example Black and Henderson (1999), Glaeser and Mare (2001), Wheeler (2001), and Hendricks (2011)).

The mechanisms for ability sorting that have been investigated are regional in nature. Either they involve migration of high quality labor to a dense region or they

involve greater or faster acquisition of human capital by workers in denser regions. Some theories posit that dense urban areas facilitate the acquisition and transmission of knowledge. For example, Glaeser (1999) develops a model of learning based on interactions. Because interactions are more common and less costly in dense areas, workers in these areas will obtain more human capital. Other theories rely on migration to enhance the average level of human capital in cities. In addition to affecting the native population, higher returns to ability or lower knowledge acquisition costs may spur migration as well. Other theories posit that dense urban areas improve the functioning of labor markets through improved matching between employees and employers such as Helsley and Strange (1990) and Kim (1990). Consumption motives have also been suggested. Lee (2010), in the spirit of the 'consumer city' of Glaeser et al. (2001), develops and tests a theory whereby highly productive workers migrate to urban areas in response to the availability of cultural amenities. He posits that demand for the amenities of large dense cities leads to the migration of high skill workers to those locations.

Unlike many of these theories that lead to endogeneity of labor quality, urban producer theory as developed in this dissertation is a small-scale sorting mechanism which operates within a single labor market rather than across labor markets. While one might expect a similar mechanism to operate between markets, possibly including industry sorting based on production technologies, many additional factors complicate that relationship. As such, it's biggest contribution to the literature on the measurement of agglomeration economies is to serve as a cautionary note regarding particular identification strategies.

In response to concerns that ability sorting across regions may bias estimates of the relationship between density (or other measures of agglomeration) and locational productivity, a number of investigators have begun to work with data disaggregated

at a level below that of the metropolitan area. This then allows for estimation within a single metropolitan area or the use of fixed effects estimation to sweep out the inter-regional differences in ability. Urban producer theory suggests that sorting on ability also occurs within metropolitan areas. To the extent that the agglomeration measures such as density are correlated with the sorting mechanism, rent, the concern over ability sorting remains. That density and rents are correlated is a well-established phenomenon. The standard profit-maximizing response to increasing land prices is to reduce lot sizes and increase building heights. Under this relationship, urban producer theory suggests that sorting within a metropolitan area can bias estimates of agglomeration economies. The concern is particularly problematic in measuring urbanization economies where researchers attempt to understand the contribution of density in general to agglomeration economies. The concern over localization economies, where it is only the density of related industries or occupations that enhance productivity, is of a lesser concern.

Two recent investigations highlight the different uses of finely detailed geographic information that allows for the control of ability sorting across regions. Each has dealt with concerns over ability sorting in a different fashion.

Fu (2007) estimates a Mincerian wage equation using individual level data for the Boston metropolitan area from the 1990 Census long form which contains detailed information on individuals including the census block in which they are employed. In addition to including the standard observable individual characteristics, Fu includes a number measures of the density of human capital and industrial and occupational composition. With this information, he is able to estimate the relative importance of measures of density on labor productivity, and therefore the extent of human capital externalities. Fu is well cognizant of the possibility and consequences of omitted individual characteristics such as worker quality that would generate a sorting bias

in his estimates and acknowledges the difficulty of accounting for the endogeneity.

Rosenthal and Strange (2008) use the 2000 Census Public Use Micro Sample (PUMS) to estimate the attenuation of agglomeration economies. They estimate a standard Mincerian wage equation that includes controls for worker and geographic characteristics. Because PUMS allows them to identify workplace locations within a metropolitan area, they use fixed effects estimation to control for ability sorting across metropolitan regions. Rosenthal and Strange pay particular attention to controlling for the endogeneity that occurs when high skilled workers are sorted into dense regions within metropolitan area. They introduce a unique method using instrumental variables. Using the standard reasoning behind selecting an instrument, they choose instruments that are correlated with density, but uncorrelated with the unobserved component of worker ability. They use geological information on bedrock, earthquake hazards, and landslide hazard. These measures are correlated with density. For example, in the New York MSA the tallest buildings, and hence the highest densities, are located in areas with easy access to solid bedrock. This generates the correlation between the included and excluded instruments. With this method they are able to estimate the effect of density on wages as well as the rate at which the effect attenuates over distance.

Urban producer theory posits an alternative sorting mechanism. Rather than high ability individuals being sorted into dense regions, the high ability individuals are sorted into high rent areas. In order for the method of Rosenthal and Strange to work, it must be the case that rents are not correlated with the the geological information. Of course we expect the price of land to vary according to the revenue it can generate and therefore land less suited for building will generally have a lower price. However the relevant price in urban producer theory is the floor area rent, not land rent. We generally expect that prices in a market will equalize based on

demand and supply factors and prices of a product will not vary depending upon the costs of individual producers. A partial answer to this issue regarding the correlation between geological characteristics and floor area rents can be obtained by examining the relationship between floor area rents and building height in Manhattan, where varying geological characteristics limit building height. Fuerst (2008) estimates the determinants of office rents in Manhattan using data from CoStar. Using a log-linear model of office rents, Fuerst (2008, table 2) estimates the elasticity of the floor area rent with respect to the number of stories in a building to vary between 0.104 and 0.215.[6] This is a relatively mild price effect though without further investigation it is not possible to determine if geological features are too weak of instruments for use in addressing the endogeneity of labor quality due to floor area rents. Furthermore, in areas outside of the densest cities such as New York, it is unclear that the geological features are a binding constraint on building height. Consequently, there is an issue whether the instruments of Rosenthal and Strange are adequate for addressing the ability sorting due to rent.

4.3 Summary

Urban producer theory provides a note of caution when using the spatial variation in wages as an identification strategy. When substantial variation in floor area rents is expected across a region, urban producer theory indicates that there will also be a substantial amount of sorting of labor based on its productivity or ability. If the explanatory variables such as the distance from the central city or density are correlated with floor areas rents, then the relation between distance or density and any dependent variable that is affected by worker productivity will be subject to omitted

[6]Table 3.11 indicates that there is not a statistically significant relationship between building height and floor area rent in Cleveland.

variable bias. By relating the variation of input quality to rents, urban producer theory provides guidance on how to eliminate omitted variable bias. Estimation should include measures of space rent. Unfortunately this may create further problems because space rent may not be exogenous in the equation of interest, i.e. in a wage or earnings equation. This method requires an excluded instrument that is correlated with floor area rent and can be excluded from the wage equation. Attempting to solve this issue by simply including rent as an explanatory variable will introduce simultaneity bias because wages and rents are jointly determined.

References

Abadie, A. and S. Dermisi (2008). Is terrorism eroding agglomeration economies in central business districts? Lessons from the office real estate market in downtown Chicago. *Journal of Urban Economics 64*, 451–463.

Abel, J. R., I. Dey, and T. M. Gabe (2012). Productivity and the density of human capital. *Journal of Regional Science 52*, 562–586.

Almazan, A., A. De Motta, and S. Titman (2007). Firm location and the creation and utilization of human capital. *Review of Economic Studies 74*, 1305–1327.

Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press.

Andersson, F., S. Burgess, and J. I. Lane (2007). Cities, matching and the productivity gains of agglomeration. *Journal of Urban Economics 61*, 112–128.

Angrist, J. D. and J.-S. Pischke (2009). *Mostly Harmless Econometrics*. Princeton University Press.

Badenhausen, K. (2011). Law school graduates do not make $160,000. Available at `http://www.forbes.com/sites/kurtbadenhausen/2011/03/23/law-school-graduates-do-not-make-160000/`. (last accessed October 20, 2012).

Berechman, J. and R. E. Paaswell (2005). Evaluation, prioritization and selection of transportation investment projects in new york city. *Transportation 32*, 223–249.

Black, D. and J. V. Henderson (1999). A theory of urban growth. *Journal of Political Economy 107*, 252–284.

Blackorby, C. and R. R. Russell (1981). The Morishima elasticity of substitution: Symmetry, constancy, separability, and its relationship to the Hicks and Allen elasticities. *Review of Economic Studies 48*, 147–158.

Blackorby, C. and R. R. Russell (1989). Will the real elasticity of substitution please stand up? (a comparison of the Allen/Uzawa and Morishima elasticities). *American Economic Review 79*, 882–888.

Boardman, A. E. and L. B. Lave (1977). Highway congestion and congestion tolls. *Journal of Urban Economics 4*, 340–359.

Bollinger, C. R., K. R. Ihlanfeldt, and D. R. Bowes (1998). Spatial variation in office rents within the Atlanta region. *Urban Studies 35*, 1097–1118.

Boylan, R. T. and C. X. Long (2005). Salaries, plea rates, and the career objectives of federal prosecutors. *Journal of Law and Economics 48*, 627–651.

Brockett, P. L., W. W. Cooper, H. Deng, L. L. Golden, and T. W. Ruefli (2004). Using DEA to identify and manage congestion. *Journal of Productivity Analysis 22*, 207–226.

Brueckner, J. K. (1987). The structure of urban equilibria: A unified treatment of the Muth-Mills model. In E. Mills (Ed.), *Handbook of Regional and Urban Economics*, Volume II, Chapter 20, pp. 821–845. Elsevier.

Byrnes, P., R. Färe, and S. Grosskopf (1984). Measuring productive efficiency: An application to Illinois strip mines. *Management Science 30*, 671–681.

Caron, P. L. and R. Gely (2006). Dead poets and academic progenitors: The next generation of law school rankings. *Indiana Law Journal 81*, 1–12.

Cleveland, W. S., S. J. Devlin, and E. Grosse (1988). Regression by local fitting: Methods, properties, and computational algorithms. *Journal of Econometrics 37*, 87–114.

Cloud, M. and G. B. Shepherd (2012). Law deans in jail. Legal Studies Research Paper No. 12-199, Emory University School of Law.

Colwell, P. and H. J. Munneke (2009). Directional land value gradients. *Journal of Real Estate Finance and Economics 39*, 1–23.

Combes, P.-P., G. Duranton, and L. Gobillon (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics 63*(2), 723–742.

Combes, P.-P., G. Duranton, and L. Gobillon (2011). The identification of agglomeration economies. *Journal of Economic Geography 11*, 253–266.

Coulson, N. E. (1991). Really useful tests of the monocentric model. *Land Economics 67*, 299–307.

Davis, D. R. and D. E. Weinstein (2008). A search for multiple equilibria in urban industrial structure. *Journal of Regional Science 48*, 29–65.

Dunse, N. and C. Jones (1998). A hedonic price model of office rents. *Journal of Property Valuation and Investment 16*, 297–312.

Eberts, R. W. (1981). An empirical investigation of intraurban wage gradients. *Journal of Urban Economics 10*, 50–60.

Eichholtz, P., N. Kok, and J. M. Quigley (2010). Doing well by doing good? Green office buildings. *American Economic Review 100*, 2492–2509.

Färe, R., S. Grosskopf, and B. Y. Yoon (1982). A theoretical and empirical analysis of the highway speed-volume relationship. *Journal of Urban Economics 12*, 115–121.

Färe, R. and L. Jansson (1975). On VES and WDI production functions. *International Economic Review 16*, 745–750.

Färe, R. and L. Svensson (1980). Congestion of production factors. *Econometrica 48*(7), 1745–1753.

Flegg, A. and D. Allen (2009). Congestion in the Chinese automobile and textile industries revisited. *Socio-Economic Planning Sciences 43*, 177–191.

Fu, S. (2007). Smart café cities: Testing human capital externalities in the Boston metropolitan area. *Journal of Urban Economics 61*, 86–111.

Fu, S. and S. L. Ross (2013). Wage premia in emploment clusters: How important is worker heterogeneity? *Journal of Labor Economics 31*, 271–304.

Fuerst, F. (2008). Office rent determinants: a hedonic panel analysis. Real Estate and Planning Working Paper rep-wp2008-12, Henley Business School, University of Reading.

Fuerst, F. and P. McAllister (2011). Green noise or green value? Measuring the effects of environmental certification on office values. *Real Estate Economics 39*, 45–69.

Garicano, L. and T. N. Hubbard (2005). Heirarchical sorting and learning costs: Theory and evidence from the law. *Journal of Economic Behavior and Organization 58*, 349–369.

Glaeser, E. and D. Mare (2001, April). Cities and skills. *Journal of Labor Economics 19*(2), 316–342.

Glaeser, E. L. (1999). Learning in cities. *Journal of Urban Economics 46*, 254–277.

Glaeser, E. L., J. Kolko, and A. Saiz (2001). Consumer city. *Journal of Economic Geography 1*, 27–50.

Greene, W. H. and D. A. Hensher (2010). *Modeling Ordered Choices: A Primer*. Cambridge University Press.

Hamermesh, D. S. (1987). The demand for labor in the long run. In O. Ashenfelter and R. Layard (Eds.), *Handbook of Labor Economics*, Volume 1, Chapter 8, pp. 429–471. Elsevier.

Haughwout, A., J. Orr, and D. Bedoll (2008). The price of land in the New York metropolitan area. *Federal Reserve Bank of New York Current Issues in Economics and Finance 14*(3), 1–7.

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

Hendricks, A. M. (1989). Hospital wage gradients within U.S. urban areas. *Journal of Health Economics 8*, 233–246.

Hendricks, L. (2011). The skill composition of U.S. cities. *International Economic Review 52*, 1–32.

Hole, A. R. (2006). Calculating Murphy-Topel variance estimates in Stata: A simplified procedure. *Stata Journal 6*, 521–529.

Ihlanfeldt, K. R. (1992). Intraurban wage gradients: evidence by race, gender, occupational class, and sector. *Journal of Urban Economics 32*(1), 70–91.

Inman, R. P. (1978). A generalized congestion function for highway travel. *Journal of Urban Economics 5*, 21–34.

Johnson, S. and G. C. Rausser (1971). Effects of misspecifications of log-linear functions when sample values are zero or negative. *American Journal of Agricultural Economics 53*, 120–124.

Kim, S. (1990). Labor heterogeneity, wage bargaining, and agglomeration economies. *Journal of Urban Economics 28*, 160–177.

Laband, D. N. and B. F. Lentz (1993). Is there sex discrimination in the legal profession? Further evidence on tangible and intangible margins. *Journal of Human Resources 28*, 230–258.

Lee, S. (2010). Ability sorting and consumer city. *Journal of Urban Economics 68*, 20–33.

Leigh, P. J. (1986). Are compensating wages paid for time spent commuting? *Applied Economics 18*, 1203–1213.

Limehouse, F. F. and R. E. McCormick (2011). Impacts of central business district location: A hedonic analysis of legal service establishments. Working Paper CES 11-21, Center for Economic Studies, U.S. Census Bureau.

Madden, J. F. (1985). Urban wage gradients: Empirical evidence. *Journal of Urban Economics 18*, 291–301.

Madden, P. (1986). *Concavity and Optimization in Microeconomics*. Basil Blackwell.

McMillen, D. P. (1996). One hundred fifty years of land values in Chicago: A nonparametric approach. *Journal of Urban Economics 40*, 100–124.

McMillen, D. P. and C. L. Redfearn (2010). Estimation and hypothesis testing for nonparametric hedonic housing price functions. *Journal of Regional Science 50*, 712–733.

McMillen, D. P. and L. D. Singell Jr. (1992). Work location, residence location, and the intraurban wage gradient. *Journal of Urban Economics 32*, 195–213.

Meese, R. and N. Wallace (1991). Nonparametric estimation of dynamic hedonic price models and the construction of residential housing price indices. *Journal of the American Real Estate and Urban Economics Association 19*, 308–332.

Mills, E. S. (1972). *Studies in the Structure of the Urban Economy*. Johns Hopkins Press.

Mion, G. and P. Naticchioni (2009, February). The spatial sorting and matching of skills and firms. *Canadian Journal of Economics 42*(1), 28–55.

Moretti, E. (2011). Local labor markets. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 4, Chapter 14, pp. 1237–1313. Elsevier.

Murphy, K. M. and R. H. Topel (1985). Estimation and inference in two-step econometric models. *Journal of Business and Economic Statistics 3*, 370–379.

Muth, R. F. (1969). *Cities and Housing: The Spatial Pattern of Urban Residential Land Use*. University of Chicago Press.

Raasch, J. E. (2007). Making partner - or not: Is it in, up or over in the 21st century? *Law Practice 33*, 32–36.

Romer, P. M. (1990). Endogeneous technological change. *Journal of Political Economy 98*, S71–102.

Rosenthal, S. S. and W. C. Strange (2008). The attenuation of human capital spillovers. *Journal of Urban Economics 64*, 373–389.

Shoup, D. C. (1995). An opportunity to reduce minimum parking requirements. *Journal of the American Planning Association 61*, 14–28.

Sivitanidou, R. (1995). Urban spatial variation in office-commercial rents: The role of spatial amenities and commercial zoning. *Journal of Urban Economics 38*, 23–49.

Söderberg, B. and C. Janssen (2001). Estimating distance gradients for apartment properties. *Urban Studies 38*, 61–79.

Spurr, S. J. (1990). Sex discrimination in the legal profession: A study of promotion. *Industrial and Labor Relations Review 43*, 406–417.

Tam, T. (1997). What's in a name: The earnings effect of american law school prestige. *EurAmerica 27*, 1–36.

Timothy, D. and W. Wheaton (2001). Intra-urban wage variation, employment location, and commuting times. *Journal of Urban Economics 50*(2), 338–366.

Waddell, P., A. Borning, M. Noth, N. Freier, M. Becke, and G. Ulfarsson (2003). Urbansim: A simulation system for land use and transportation. *Networks and Spatial Economics 3*, 43–67.

Wheaton, W. C., M. S. Baranski, and C. A. Templeton (2009). 100 years of commercial real estate prices in Manhattan. *Real Estate Economics 37*, 69–83.

Wheeler, C. H. (2001). Search, sorting, and urban agglomeration. *Journal of Labor Economics 19*, 879–899.

Appendices

A Proofs

Convavity of the Urban Production Function with Separable Input Quality

$$
\begin{aligned}
Q &= [\alpha N^{\rho} + \beta H^{\rho} + \phi\,(S - \gamma N)^{\rho}]^{\frac{1}{\rho}} \quad \text{if } (S - \gamma N) > 0 \\
Q &= 0 \quad \text{if } (S - \gamma N) \leq 0 \\
\text{where } \rho &\in (-\infty, 1) \\
&\quad \alpha, \beta, \phi, \gamma > 0
\end{aligned}
$$

Express in terms of composite functions

$$
\begin{aligned}
Q(S,N,H) &= e^{f(u)} \\
f(u) &= \frac{1}{\rho}\ln\left(h(u)\right) \\
h(S,N,H) &= \alpha j(S,N,H) + \beta k\left(S,N,H\right) + \phi l(S,N,H) \\
j\left(S,N,H\right) &= N^{\rho} \\
k\left(S,N,H\right) &= H^{\rho} \\
l\left(S,N,H\right) &= (S-\gamma N)^{\rho}
\end{aligned}
$$

The strategy is to show the properties of $j(S,N,H)$, $k(S,N,H)$, and $l\left(S,N,H\right)$ then rely upon theorems regarding transformations of convex and concave function. This is based upon the presentation in Madden (1986). The following theorems are used in the proof.

Theorem 1. *If $r\left(\mathbf{x}\right)$ and $s\left(\mathbf{x}\right)$ are concave (convex) and a and b are non-negative, then $a \cdot r(\mathbf{x}) + b \cdot s(\mathbf{x})$ is concave (convex). (Theorem 13.5 in Madden (1986))*

Theorem 2. *If $r\left(\mathbf{x}\right)$ is concave (convex) and $s\left(u\right)$ is concave (convex) and monotone increasing, then $U\left(u\right) = s\left(r\left(\mathbf{x}\right)\right)$ is concave (convex). (Theorem 13.6 in Madden (1986))*

Theorem 3. *If $r\left(\mathbf{x}\right)$ is concave (convex), then $r\left(\mathbf{x}\right)$ is quasi-concave (quasi-convex).*

Theorem 4. *If $r\left(\mathbf{x}\right)$ is quasi-concave (quasi-convex) and $s\left(u\right)$ is monotone increasing, then $U\left(u\right) = s\left(r\left(\mathbf{x}\right)\right)$ is quasi-concave (quasi-convex). (Theorem 13.8 in Madden (1986))*

Theorem 5. *If $r\left(\mathbf{x}\right)$ is C^1, linearly homogeneous, and positively valued, then $r\left(\mathbf{x}\right)$ is concave if and only if $r\left(\mathbf{x}\right)$ is quasi-concave. (Theorem 13.3 in Madden (1986))*

Theorem 6. *Suppose* $r(\mathbf{x}) = k \cdot s(\mathbf{x})$. *If* $s(\mathbf{x})$ *is concave (quasi-concave),* $r(\mathbf{x})$ *will be concave (quasi-convcave) if* $k > 0$ *and* $r(\mathbf{x})$ *will be convex (quasi-convex) if* $k < 0$. *If* $s(\mathbf{x})$ *is convex (quasi-convex),* $r(\mathbf{x})$ *will be convex (quasi-convex) if* $k > 0$ *and* $r(\mathbf{x})$ *will be concave (quasi-concave) if* $k < 0$. *(Theorems 13.4 and 13.7 in Madden (1986))*

Properties of $j(S, N, H)$

Determine quadratic form of $j(S, N, H)$ by examining the Hessian matrix.

$$\begin{aligned}
\frac{\partial j(S,N,H)}{\partial N} &= \rho N^{\rho-1} \\
\frac{\partial j(S,N,H)}{\partial H} &= 0 \\
\frac{\partial j(S,N,H)}{\partial S} &= 0 \\
\frac{\partial^2 j(S,N,H)}{\partial N^2} &= \rho(\rho-1) N^{\rho-2} \\
\frac{\partial^2 j(S,N,H)}{\partial H^2} &= 0 \\
\frac{\partial^2 j(S,N,H)}{\partial S^2} &= 0 \\
\frac{\partial^2 j(S,N,H)}{\partial N \partial H} &= 0 \\
\frac{\partial^2 j(S,N,H)}{\partial N \partial S} &= 0 \\
\frac{\partial^2 j(S,N,H)}{\partial S \partial H} &= 0
\end{aligned}$$

The Hessian of $j(S, N, H)$ is

$$j'' = \begin{bmatrix} \rho(\rho-1)N^{\rho-2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The Hessian is diagonal and therefore we need only check the signs of the diagonal elements to determine the definiteness of the Hessian. The diagonal elements are $\rho(\rho-1)N^{\rho-2}$, 0, and 0. The sign of the first diagonal element will depend on the value of ρ.

$$\begin{aligned} \rho(\rho-1)N^{\rho-2} &\geq 0 \ \forall \rho \leq 0 \\ \rho(\rho-1)N^{\rho-2} &\leq 0 \ \forall \ 0 \leq \rho < 1 \end{aligned}$$

Therefore, when $\rho \leq 0$ the Hessian matrix is positive semi-definite and consequently $j(S,N,H)$ is convex. When $0 \leq \rho < 1$ the Hessian matrix is negative semi-definite and consequently $j(S,N,H)$ is concave.

Properties of $k(S,N,H)$

Determine quadratic form of $k(S,N,H)$ by examining the Hessian matrix

$$
\begin{aligned}
\frac{\partial k\left(S,N,H\right)}{\partial N} &= 0 \\
\frac{\partial k\left(S,N,H\right)}{\partial H} &= \rho H^{\rho-1} \\
\frac{\partial k\left(S,N,H\right)}{\partial S} &= 0 \\
\frac{\partial^2 k\left(S,N,H\right)}{\partial N^2} &= 0 \\
\frac{\partial^2 k\left(S,N,H\right)}{\partial H^2} &= \rho\left(\rho-1\right)H^{\rho-2} \\
\frac{\partial^2 k\left(S,N,H\right)}{\partial S^2} &= 0 \\
\frac{\partial^2 k\left(S,N,H\right)}{\partial N \partial H} &= 0 \\
\frac{\partial^2 k\left(S,N,H\right)}{\partial N \partial S} &= 0 \\
\frac{\partial^2 k\left(S,N,H\right)}{\partial S \partial H} &= 0
\end{aligned}
$$

The Hessian of $k\left(S,N,H\right)$ is

$$
k'' = \begin{bmatrix}
0 & 0 & 0 \\
0 & \rho\left(\rho-1\right)H^{\rho-2} & 0 \\
0 & 0 & 0
\end{bmatrix}
$$

The Hessian is diagonal, therefore I need only check the signs of the diagonal elements to determine the definiteness of the Hessian. The diagonal elements are 0, $\rho\left(\rho-1\right)N^{\rho-2}$, and 0. The sign of the second diagonal element will depend on the value of ρ.

$$\begin{aligned}
\rho(\rho-1)H^{\rho-2} &\geq 0 \;\forall \rho \leq 0 \\
\rho(\rho-1)H^{\rho-2} &\leq 0 \;\forall\, 0 \leq \rho < 1
\end{aligned}$$

Therefore, when $\rho \leq 0$ the Hessian matrix is positive semi-definite and consequently $k(S,N,H)$ is convex. When $0 \leq \rho < 1$ the Hessian matrix is negative semi-definite and consequently $k(S,N,H)$ is concave.

Properties of $l(S,N)$

Determine quadratic form of $l(S,N)$ by examining the Hessian matrix

$$\begin{aligned}
\frac{\partial l(S,N,H)}{\partial N} &= -\rho\gamma(S-\gamma N)^{\rho-1} \\
\frac{\partial l(S,N,H)}{\partial H} &= 0 \\
\frac{\partial l(S,N,H)}{\partial S} &= \rho(S-\gamma N)^{\rho-1} \\
\frac{\partial^2 l(S,N,H)}{\partial N^2} &= \rho(\rho-1)\gamma^2(S-\gamma N)^{\rho-2} \\
\frac{\partial^2 l(S,N,H)}{\partial H^2} &= 0 \\
\frac{\partial^2 l(S,N,H)}{\partial S^2} &= \rho(\rho-1)(S-\gamma N)^{\rho-2} \\
\frac{\partial^2 l(S,N,H)}{\partial N \partial H} &= 0 \\
\frac{\partial^2 l(S,N,H)}{\partial N \partial S} &= -\rho(\rho-1)\gamma(S-\gamma N)^{\rho-2} \\
\frac{\partial^2 l(S,N,H)}{\partial S \partial H} &= 0
\end{aligned}$$

The Hessian of $l(S,N,H)$ is

$$l'' = \begin{bmatrix} \rho(\rho-1)\gamma^2(S-\gamma N)^{\rho-2} & 0 & -\rho(\rho-1)\gamma(S-\gamma N)^{\rho-2} \\ 0 & 0 & 0 \\ -\rho(\rho-1)\gamma(S-\gamma N)^{\rho-2} & 0 & \rho(\rho-1)(S-\gamma N)^{\rho-2} \end{bmatrix}$$

The Hessian is not diagonal and therefore I must check the signs of the all of the principal minors.

The first-order principal minors of the Hessian are $\rho(\rho-1)\gamma^2(S-\gamma N)^{\rho-2}$, 0, and $\rho(\rho-1)(S-\gamma N)^{\rho-2}$. The signs of several terms depend on the value of ρ.

$$\begin{aligned} \rho(\rho-1)\gamma^2(S-\gamma N)^{\rho-2} &\geq 0 \ \forall \rho \leq 0 \\ \rho(\rho-1)\gamma^2(S-\gamma N)^{\rho-2} &\leq 0 \ \forall \ 0 \leq \rho < 1 \\ \rho(\rho-1)(S-\gamma N)^{\rho-2} &\geq 0 \ \forall \rho \leq 0 \\ \rho(\rho-1)(S-\gamma N)^{\rho-2} &\leq 0 \ \forall \ 0 \leq \rho < 1 \end{aligned}$$

When $\rho \leq 0$, the first-order principal minors will be non-negative. When $\rho \in [\,0, 1\,)$ the first-order principal minors are non-positive.

The second-order principal minors of the Hessian are:

$$\begin{vmatrix} \rho(\rho-1)\gamma^2(S-\gamma N)^{\rho-2} & 0 \\ 0 & 0 \end{vmatrix} = \left(\rho(\rho-1)\gamma^2(S-\gamma N)^{\rho-2}\right)(0) - (0)(0) = 0$$

$$\begin{vmatrix} 0 & 0 \\ 0 & \rho(\rho-1)(S-\gamma N)^{\rho-2} \end{vmatrix} = (0)\left(\rho(\rho-1)(S-\gamma N)^{\rho-2}\right) - (0)(0) = 0$$

$$\begin{vmatrix} \rho(\rho-1)\gamma^2(S-\gamma N)^{\rho-2} & -\rho(\rho-1)\gamma(S-\gamma N)^{\rho-2} \\ -\rho(\rho-1)\gamma(S-\gamma N)^{\rho-2} & \rho(\rho-1)(S-\gamma N)^{\rho-2} \end{vmatrix} =$$

$$\left[\rho(\rho-1)\gamma^2(S-\gamma N)^{\rho-2}\right]\left[\rho(\rho-1)(S-\gamma N)^{\rho-2}\right] -$$

$$\left[-\rho(\rho-1)\gamma(S-\gamma N)^{\rho-2}\right]\left[-\rho(\rho-1)\gamma(S-\gamma N)^{\rho-2}\right] =$$

$$\rho^2(\rho-1)^2\gamma^2(S-\gamma N)^{2\rho-4} - \rho^2(\rho-1)^2\gamma^2(S-\gamma N)^{2\rho-4} = 0$$

The third-order principal minor of the Hessian of $l(K,N,H)$ is

$$\begin{vmatrix} \rho(\rho-1)\gamma^2(S-\gamma N)^{\rho-2} & 0 & -\rho(\rho-1)\gamma(S-\gamma N)^{\rho-2} \\ 0 & 0 & 0 \\ -\rho(\rho-1)\gamma(S-\gamma N)^{\rho-2} & 0 & \rho(\rho-1)(S-\gamma N)^{\rho-2} \end{vmatrix} =$$

$$\left(\rho(\rho-1)\gamma^2(S-\gamma N)^{\rho-2}\right)\begin{vmatrix} 0 & 0 \\ 0 & \rho(\rho-1)(S-\gamma N)^{\rho-2} \end{vmatrix} -$$

$$(0)\begin{vmatrix} 0 & 0 \\ -\rho(\rho-1)\gamma(S-\gamma N)^{\rho-2} & \rho(\rho-1)(S-\gamma N)^{\rho-2} \end{vmatrix} +$$

$$\left(-\rho(\rho-1)\gamma(S-\gamma N)^{\rho-2}\right)\begin{vmatrix} 0 & 0 \\ -\rho(\rho-1)\gamma(S-\gamma N)^{\rho-2} & 0 \end{vmatrix} =$$

$$0 - 0 + 0 = 0$$

When $\rho \leq 0$, the principal minors are all non-negative indicating that the Hessian is positive semi-definite and therefore $l(S,N,H)$ is a convex function when $\rho \leq 0$.

When $0 \leq \rho < 1$, the principal minors alternate in sign with the first-order principal minors being non-positive and the second-order principal minor being non-

negative. This indicates that the Hessian is negative semi-definite and therefore $l(S,N,H)$ is concave when $0 \leq \rho < 1$.

Properties of $Q(S,N,H)$ when $0 \leq \rho < 1$

$j(S,N,H), k(S,N,H), l(S,N,H)$ are concave with $0 \leq \rho < 1$ and $h(S,N,H) = \alpha j(S,N,H) + \beta k(S,N,H) + \phi l(S,N,H)$ with $\alpha, \beta, \phi > 0$. By theorem 1, $h(S,N,H) = \alpha N^{\rho} + \beta H^{\rho} + \phi (S - \gamma N)^{\rho}$ is concave when $0 \leq \rho < 1$.

$f(u) = \frac{1}{\rho} \ln(u)$ which is concave and monotone increasing in u when $0 \leq \rho \leq 1$. By theorem 2, $f(h(S,N,H)) = \frac{1}{\rho} \ln(\alpha N^{\rho} + \beta H^{\rho} + \phi (S - \gamma N)^{\rho})$ is concave.

By theorem 3 we have that $f(u)$ is also quasi-concave.

$Q(u) = e^{f(u)}$ where $Q(u)$ is monotone increasing in u, though it is not concave. By theorem 4, $Q(u) = e^{f(u)} = e^{\frac{1}{\rho} \ln(\alpha N^{\rho} + \beta H^{\rho} + \phi (S - \gamma N)^{\rho})} = [\alpha N^{\rho} + \beta H^{\rho} + \phi (S - \gamma N)^{\rho}]^{\frac{1}{\rho}}$ is quasi-concave.

$Q(S,N,H)$ is homogeneous of degree one. Therefore by theorem 5, $Q(S,N,H)$ is concave when $0 \leq \rho < 1$.

Properties of $Q(S,N,H)$ when $\rho < 0$

It is the case that $j(S,N,H), k(S,N,H), l(S,N,H)$ are convex with $\rho < 0$ and $h(S,N,H) = \alpha j(S,N,H) + \beta k(S,N,H) + \phi l(S,N,H)$ with $\alpha, \beta, \phi > 0$. By theorem 1, $h(S,N,H) = \alpha N^{\rho} + \beta H^{\rho} + \phi (S - \gamma N)^{\rho}$ is convex when $\rho < 0$.

By theorem 3 $h(S,N,H)$ is also quasi-convex.

$\ln(u)$ is monotonically increasing. By theorem 4 $\ln(h(S,N,H))$ is quasi-convex.

$f(u) = \frac{1}{\rho} \ln(u)$ with $\frac{1}{\rho} < 0$ when $\rho < 0$. By theorem 6 $\frac{1}{\rho} \ln(h(S,N,H))$ is quasi-concave when $\rho < 0$.

$f(u) = \frac{1}{\rho} \ln(u)$ is concave and monotone increasing in u when $\rho \geq 0$. By theorem 2, $f(h(S,N,H)) = \frac{1}{\rho} \ln(\alpha N^{\rho} + \beta H^{\rho} + \phi (S - \gamma N)^{\rho})$ is concave.

By theorem 3 $f(u)$ is also quasi-concave.

$Q(u) = e^{f(u)}$ where $Q(u)$ is monotone increasing in u, though it is not concave. By theorem 4, $Q(u) = e^{f(u)} = e^{\frac{1}{\rho}\ln(\alpha N^\rho + \beta H^\rho + \phi(S-\gamma N)^\rho)} = [\alpha N^\rho + \beta H^\rho + \phi(S-\gamma N)^\rho]^{\frac{1}{\rho}}$ is quasi-concave.

$Q(S, N, H)$ is homogeneous of degree one. Therefore by theorem 5, $Q(S, N, H)$ is concave when $\rho < 0$.

■

Sign of $\frac{\partial^2 h}{\partial r \partial \gamma}$ in the Separable Quality Model

$$\frac{\partial^2 h}{\partial r \partial \gamma} = \frac{1}{1-\rho}\left(\frac{\beta}{\alpha g}\right)^{\frac{1}{1-\rho}}(w+\gamma r)^{\frac{\rho}{1-\rho}}\left[1+\left(\frac{\rho}{1-\rho}\right)\left(\frac{\gamma r}{w+\gamma r}\right)\right] \tag{4.1}$$

The term in the brackets is clearly non-negative for $\rho \in [0, 1)$. However, when $\rho < 0$ the term in the brackets must be examined more closely. Note that $\frac{\gamma r}{w+\gamma r} < 1$ for all positive values. The term $\frac{\rho}{1-\rho} \in (-1, 0)$ when $\rho < 0$. Therefore $\left(\frac{\rho}{1-\rho}\right)\left(\frac{\gamma r}{w+\gamma r}\right) > -1$ and $\frac{\partial h}{\partial r \partial \gamma} > 0$

■

Sign of $\frac{\partial \epsilon_{hr}}{\partial \gamma}$ in the Separable Quality Model

$$\epsilon_{hr} = \frac{\gamma r}{(1-\rho)(w+\gamma r)} \tag{4.2}$$

and therefore

$$\frac{\partial \epsilon_{hr}}{\partial \gamma} = \frac{1}{1-\rho} r (w+\gamma r)^{-1}\left[1 - \frac{\gamma r}{w+\gamma r}\right] \tag{4.3}$$

The term in the brackets is less than 1 when $\gamma, r, w > 0$ and therefore $\frac{\partial \epsilon_{hr}}{\partial \gamma} > 0$ for admissible values of ρ.

■

Sign of $\frac{\partial^2 n}{\partial r \partial \gamma_L}$ in the Embedded Quality Model

Begin by rewriting $\frac{\partial n}{\partial r}$ in a more tractable form for taking the derivative with respect to γ_L

$$\frac{\partial n}{\partial r} = (1-\rho)^{-1} \left(\frac{\alpha_H}{\alpha_L}\right)^{\frac{1}{1-\rho}} (w_L + \gamma_L r)^{\frac{\rho}{1-\rho}} (w_H + \gamma_H r)^{\frac{\rho-2}{1-\rho}} (\gamma_L w_H - \gamma_H w_L)$$

Therefore,

$$\begin{aligned}\frac{\partial n}{\partial r \partial \gamma_L} = &(1-\rho)^{-1} \left(\frac{\alpha_H}{\alpha_L}\right)^{\frac{1}{1-\rho}} (w_H + \gamma_H r)^{\frac{\rho-2}{1-\rho}} \\ &\left[\frac{\rho}{1-\rho} r (w_L + \gamma_L r)^{\frac{\rho}{1-\rho}-1} (\gamma_L w_H - \gamma_H w_L) + w_H (w_L + \gamma_L r)^{\frac{\rho}{1-\rho}}\right]\end{aligned}$$

In the above equation it is clear that for $\rho > 0$ the expression will be positive under the continuing assumption that $\gamma_L w_H - \gamma_H w_L > 0$. However the result is ambiguous when $\rho < 0$ because the first term in the brackets is negative while the second term is positive. Collecting terms and simplifying the contents in the brackets will clarify the results.

$$
\begin{aligned}
\frac{\partial n}{\partial r \partial \gamma_L} &= (1-\rho)^{-1} \left(\frac{\alpha_H}{\alpha_L}\right)^{\frac{1}{1-\rho}} (w_H + \gamma_H r)^{\frac{\rho-2}{1-\rho}} (w_L + \gamma_L r)^{\frac{\rho}{1-\rho}} \\
&\quad \left[\frac{\rho}{1-\rho} r (w_L + \gamma_L r)^{-1} (\gamma_L w_H - \gamma_H w_L) + w_H\right] \\
&= (1-\rho)^{-1} \left(\frac{\alpha_H}{\alpha_L}\right)^{\frac{1}{1-\rho}} (w_H + \gamma_H r)^{\frac{\rho-2}{1-\rho}} (w_L + \gamma_L r)^{\frac{\rho}{1-\rho}} \\
&\quad \left[\frac{r\rho (\gamma_L w_H - \gamma_H w_L)}{(1-\rho)(w_L + \gamma_L r)} + \frac{w_H (1-\rho)(w_L + \gamma_L r)}{(1-\rho)(w_L + \gamma_L r)}\right] \\
&= (1-\rho)^{-1} \left(\frac{\alpha_H}{\alpha_L}\right)^{\frac{1}{1-\rho}} (w_H + \gamma_H r)^{\frac{\rho-2}{1-\rho}} (w_L + \gamma_L r)^{\frac{\rho}{1-\rho}} \\
&\quad \left[\frac{r\rho\gamma_L w_H - r\rho\gamma_H w_L}{(1-\rho)(w_L + \gamma_L r)} + \frac{w_H w_L + w_H \gamma_L r - \rho w_H w_L - \rho w_H \gamma_L r}{(1-\rho)(w_L + \gamma_L r)}\right] \\
&= (1-\rho)^{-1} \left(\frac{\alpha_H}{\alpha_L}\right)^{\frac{1}{1-\rho}} (w_H + \gamma_H r)^{\frac{\rho-2}{1-\rho}} (w_L + \gamma_L r)^{\frac{\rho}{1-\rho}} \\
&\quad \left[\frac{r\rho\gamma_L w_H - r\rho\gamma_H w_L + w_H w_L + w_H \gamma_L r - \rho w_H w_L - \rho w_H \gamma_L r}{(1-\rho)(w_L + \gamma_L r)}\right] \\
&= (1-\rho)^{-1} \left(\frac{\alpha_H}{\alpha_L}\right)^{\frac{1}{1-\rho}} (w_H + \gamma_H r)^{\frac{\rho-2}{1-\rho}} (w_L + \gamma_L r)^{\frac{\rho}{1-\rho}} \\
&\quad \left[\frac{(1-\rho) w_H w_L + r (w_H \gamma_L r - \rho \gamma_H w_L)}{(1-\rho)(w_L + \gamma_L r)}\right]
\end{aligned}
$$

The assumption that $\gamma_L w_H - \gamma_H w_L > 0$ also implies that $\gamma_L w_H - \rho\gamma_H w_L > 0$ for $\rho < 1$. Therefore all terms are positive and $\frac{\partial n}{\partial r \partial \gamma_L} > 0$.

■

Effect of r on the Elasticity of Input Quality in the Embedded Quality Model

Begin by multiplying out terms in the expression for elasticity

$$
\epsilon_{nr} = (1-\rho)^{-1} (\gamma_L w_H - \gamma_H w_L) r \left(w_L w_H + \gamma_L r w_H + w_L \gamma_H r + r^2 \gamma_L \gamma_H\right)^{-1}
$$

Therefore,

$$\frac{\partial \epsilon_{nr}}{\partial r} = (1-\rho)^{-1}(\gamma_L w_H - \gamma_H w_L)\left[(w_L+\gamma_L r)^{-1}(w_H+\gamma_H r)^{-1} - r(w_L+\gamma_L r)^{-2}(w_H+\gamma_H r)^{-2}(\gamma_L w_H + \gamma_H w_L + 2\gamma_L\gamma_H)\right]$$

$$= (1-\rho)^{-1}(\gamma_L w_H - \gamma_H w_L)(w_L+\gamma_L r)^{-1}(w_H+\gamma_H r)^{-1} \times \left[1 - \frac{\gamma_L r w_H + \gamma_H r w_L + 2r\gamma_L\gamma_H}{w_L w_H + \gamma_L r w_H + w_L\gamma_H r + r^2\gamma_L\gamma_H}\right]$$

$$= (1-\rho)^{-1}(\gamma_L w_H - \gamma_H w_L)(w_L+\gamma_L r)^{-1}(w_H+\gamma_H r)^{-1} \times \left[1 - \frac{\gamma_L r w_H + \gamma_H r w_L + 2r\gamma_L\gamma_H + r^2\gamma_L\gamma_H + w_L w_H - r^2\gamma_L\gamma_H - w_L w_H}{w_L w_H + \gamma_L r w_H + w_L\gamma_H r + r^2\gamma_L\gamma_H}\right]$$

$$= (1-\rho)^{-1}(\gamma_L w_H - \gamma_H w_L)(w_L+\gamma_L r)^{-1}(w_H+\gamma_H r)^{-1} \times \left[\frac{w_L w_H + r^2\gamma_L\gamma_H - 2r\gamma_L\gamma_H}{(w_L+\gamma_L r)(w_H+\gamma_H r)}\right]$$

All terms are positive except for $w_L w_H + r^2\gamma_L\gamma_H - 2r\gamma_L\gamma_H$ which is indeterminate. Expressing this as $w_L w_H + (r^2 - 2r)\gamma_L\gamma_H$ makes it clear that the second term will be smallest when $r = 1$. Therefore I can be assured that the elasticity is increasing in r when $w_L w_H - \gamma_L\gamma_H > 0$.

■

Sign of $\frac{\partial \epsilon_{nr}}{\partial \gamma_L}$ in the Embedded Quality Model

$$\epsilon_{nr} = \frac{r}{1-\rho}\left(\gamma_L w_H - \gamma_H w_L\right)\left(w_L + \gamma_L r\right)^{-1}\left(w_H + \gamma_H r\right)^{-1}$$

Therefore

$$\frac{\partial \epsilon_{nr}}{\partial \gamma_L} = \frac{r}{1-\rho}\left(w_H + \gamma_H r\right)^{-1}\left[w_H\left(w_L + \gamma_L r\right)^{-1} - r\left(\gamma_L w_H - \gamma_H w_L\right)\left(w_L + \gamma_L r\right)^{-2}\right]$$

$$= \frac{r}{1-\rho}\left(w_H + \gamma_H r\right)^{-1}\left(w_L + \gamma_L r\right)^{-1}\left[w_H - r\left(\gamma_L w_H - \gamma_H w_L\right)\left(w_L + \gamma_L r\right)^{-1}\right]$$

$$= \frac{r}{1-\rho}\left(w_H + \gamma_H r\right)^{-1}\left(w_L + \gamma_L r\right)^{-1}\left[\frac{w_H w_L + w_L \gamma_H r}{w_L + \gamma_L r}\right]$$

All terms are positive and therefore $\frac{\partial \epsilon_{nr}}{\partial \gamma_L} > 0$

■

Sign of $\frac{\partial \widetilde{s}}{\partial r}$ in the Embedded Quality Model

$$\widetilde{s} = \frac{\left(\frac{\phi}{r}\right)^{\frac{1}{1-\rho}}}{\left(\frac{\alpha_L}{w_L+\gamma r}\right)^{\frac{1}{1-\rho}} + \left(\frac{\alpha_H}{w_H+\gamma r}\right)^{\frac{1}{1-\rho}}} + \gamma$$

$$\frac{\partial \widetilde{s}}{\partial r} = \frac{\left[\frac{1}{\rho-1}\right]\left[\frac{\phi}{r}\right]^{\frac{1}{1-\rho}}\left[\frac{1}{r}\right]\left[\left(\frac{\alpha_L}{w_L+\gamma r}\right)^{\frac{1}{1-\rho}}+\left(\frac{\alpha_H}{w_H+\gamma r}\right)^{\frac{1}{1-\rho}}\right]}{\left[\left(\frac{\alpha_L}{w_L+\gamma r}\right)^{\frac{1}{1-\rho}}+\left(\frac{\alpha_H}{w_H+\gamma r}\right)^{\frac{1}{1-\rho}}\right]^2} -$$

$$\frac{\left[\frac{\phi}{r}\right]^{\frac{1}{1-\rho}}\left[\gamma\left(\frac{1}{\rho-1}\right)\left(\frac{\alpha_L}{w_L+\gamma r}\right)^{\frac{1}{1-\rho}}(w_L+\gamma r)^{-1}+\gamma\left(\frac{1}{\rho-1}\right)\left(\frac{\alpha_H}{w_H+\gamma r}\right)^{\frac{1}{1-\rho}}(w_H+\gamma r)^{-1}\right]}{\left[\left(\frac{\alpha_L}{w_L+\gamma r}\right)^{\frac{1}{1-\rho}}+\left(\frac{\alpha_H}{w_H+\gamma r}\right)^{\frac{1}{1-\rho}}\right]^2}$$

$$= \frac{\left[\frac{1}{\rho-1}\right]\left[\frac{\phi}{r}\right]^{\frac{1}{1-\rho}}\left[\left(\frac{1}{r}-\frac{\gamma}{w_L+\gamma r}\right)\left(\frac{\alpha_L}{w_L+\gamma r}\right)^{\frac{1}{1-\rho}}+\left(\frac{1}{r}-\frac{\gamma}{w_H+\gamma r}\right)\left(\frac{\alpha_H}{w_H+\gamma r}\right)^{\frac{1}{1-\rho}}\right]}{\left[\left(\frac{\alpha_L}{w_L+\gamma r}\right)^{\frac{1}{1-\rho}}+\left(\frac{\alpha_H}{w_H+\gamma r}\right)^{\frac{1}{1-\rho}}\right]^2}$$

$$= \frac{\left[\frac{1}{\rho-1}\right]\left[\frac{\phi}{r}\right]^{\frac{1}{1-\rho}}\left[\left(\frac{w_L}{r(w_L+\gamma r)}\right)\left(\frac{\alpha_L}{w_L+\gamma r}\right)^{\frac{1}{1-\rho}}+\left(\frac{w_H}{r(w_H+\gamma r)}\right)\left(\frac{\alpha_H}{w_H+\gamma r}\right)^{\frac{1}{1-\rho}}\right]}{\left[\left(\frac{\alpha_L}{w_L+\gamma r}\right)^{\frac{1}{1-\rho}}+\left(\frac{\alpha_H}{w_H+\gamma r}\right)^{\frac{1}{1-\rho}}\right]^2}$$

All terms are positive expect for $\frac{1}{\rho-1}$ which is negative for all admissible values of ρ. Therefore $\frac{\partial \widetilde{s}}{\partial r} < 0$.

■

B Sensitivity Analysis of Error Structure in the Rent Estimation

This appendix examines the stability of the hypothesis tests of urban producer theory under different assumptions regarding the error structure of the rent estimation procedure. Because the error terms of spaces located in the same property may be correlated, the standard correction would be to use the robust covariance matrix adjusted for clustered error terms as described by Angrist and Pischke (2009). The robust non-IID covariance matrix is calculated as

$$\left[\frac{N-1}{N-k}\right]\left[\frac{M}{M-1}\right](X'X))^{-1}\left[\sum_{m=1}^{M} u_m' u_m\right](X'X))^{-1} \tag{4.4}$$

where,

$$u_m = \sum_{j \in m}\left(y_j - X_j'\beta\right)X_j$$

$m = 1, \ldots M$ indexes the properties

$J = 1, \ldots N$ indexes the spaces

k is the number of independent variables in the rent regression

In this expression, the role for the geographic weights is unclear. In empirical models in which weights are used, such as with survey data, asymptotic theory provides guidance for incorporating the weights into procedures to obtain a consistent estimate of the covariance matrix. However, no similar work exists relating geographic weights to a consistent covariance matrix. Use of a covariance matrix that is not consistent may yield false conclusions regarding hypothesis tests of urban producer theory. In order to test the sensitivity of my results to the assumptions regarding the error structure of the rent regressions, I replicate the analysis under three alternative

specifications for the rent regressions. The first alternative, denoted LWR-nonIID-1, estimates the locally weighted least squares regression and then blindly calculates the robust covariance matrix using equation (4.4). The second alternative, denoted LWR-nonIID-2, is to estimate the locally weighted least squares regression and assume that the variance of the error terms are inversely proportional to the weights. This is a situation in which the asymptotic properties of the robust covariance matrix are known.[7] The final specification eschews the geographic weights altogether, though maintains the 30% bandwidth, and estimates the covariance matrix using equation (4.4). This specification is denoted LS-nonIID. These three alternatives can then be compared to the results of section 3.5.3, which I denote as LWR-IID.

Tables B.1 through B.4 present the results from each of the four ordered logit models under my preferred error structure, LWR-IID, as well as the three alternatives discussed here. The results clearly indicate that the hypothesis tests of urban producer theory are robust to alternative error structures in the first stage regression. While the point estimates in the tables for alternatives LWR-IID, LWR-nonIID-1, and LWR-nonIID-2 must be identical, the estimated standard errors also appear to be identical. This is a result of the rounding in the reporting. The differences in the standard errors are sufficiently minor to be lost in the rounding. For example, the standard error for the rent variable in alternatives LWR-IID, LWR-nonIID-1, and LWR-nonIID-2 of model 3 are 0.0110519, 0.0113764, and 0.011036, respectively. The z-statistics on the null hypothesis test are 5.01, 4.86, and 5.01. The resulting confidence intervals of the elasticities and probabilities of low and high quality attorneys are illustrated by figures B.1 through B.7. Naturally, the estimated coefficients of the locally weighted regressions are identical, however, the coefficients of the non-weighted least squares

[7]The robust covariance matrix in this instance is estimated by normalizing the weights to sum to N and then using that weight when constructing u_m.

regression are very similar to their weighted counterparts. The similarity of the elasticities and probabilities between weighted and unweighted rent regressions can be seen in figures B.8 and B.9.

The conclusion of these sensitivity tests are that my results are robust to alternative assumptions in regards to the error structure, as well as the choice to use geographic weights in the rent regressions. In the section that follows, I examine the robustness of my results to the selection of the bandwidth used in the rent estimation.

C Sensitivity Analysis of Bandwidth in the Rent Estimation

The locally weighted or lowess regression method only uses a portion of the data to estimate the rent regressions. In section 3.5.3 I use a bandwidth of 30%. In order to examine the sensitivity of the tests of urban producer theory, I replicate the analysis of section 3.5.3 with several other bandwidths. Tables C.1 through C.4 present the results using alternative bandwidths for each of the four versions of the ordered logit model. The results show that the tests of the urban producer hypothesis are quite consistent across bandwidths. Departures from the consistency occur at the higher bandwidths such as 65% and 85%. Figures C.1 through C.7 illustrate the elasticities and probabilities of low and high quality attorneys. What is striking in these figures is the tight grouping of the predictions that include the 30% bandwidth. This supports the conclusion that the tests of urban producer theory are robust to reasonable choices of bandwidth.

Figures

Figure 2.1: The effect of ρ on the input quality gradient

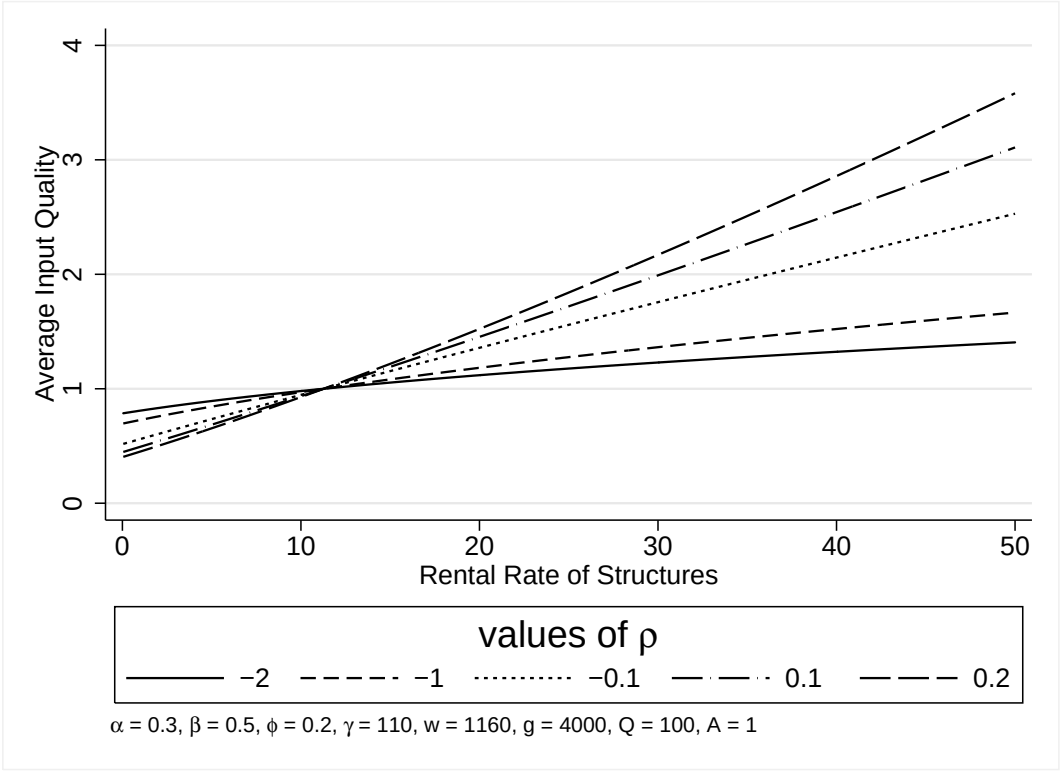


Figure 2.2: The effect of α on the input quality gradient

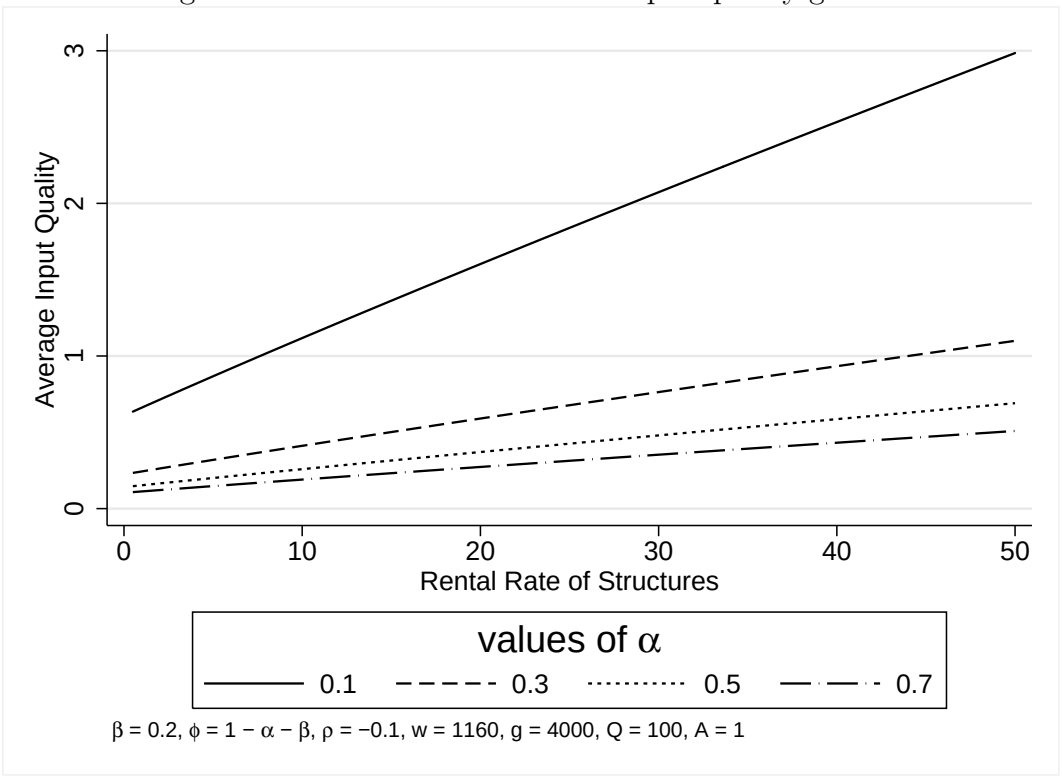


Figure 2.3: The effect of γ on the input quality gradient

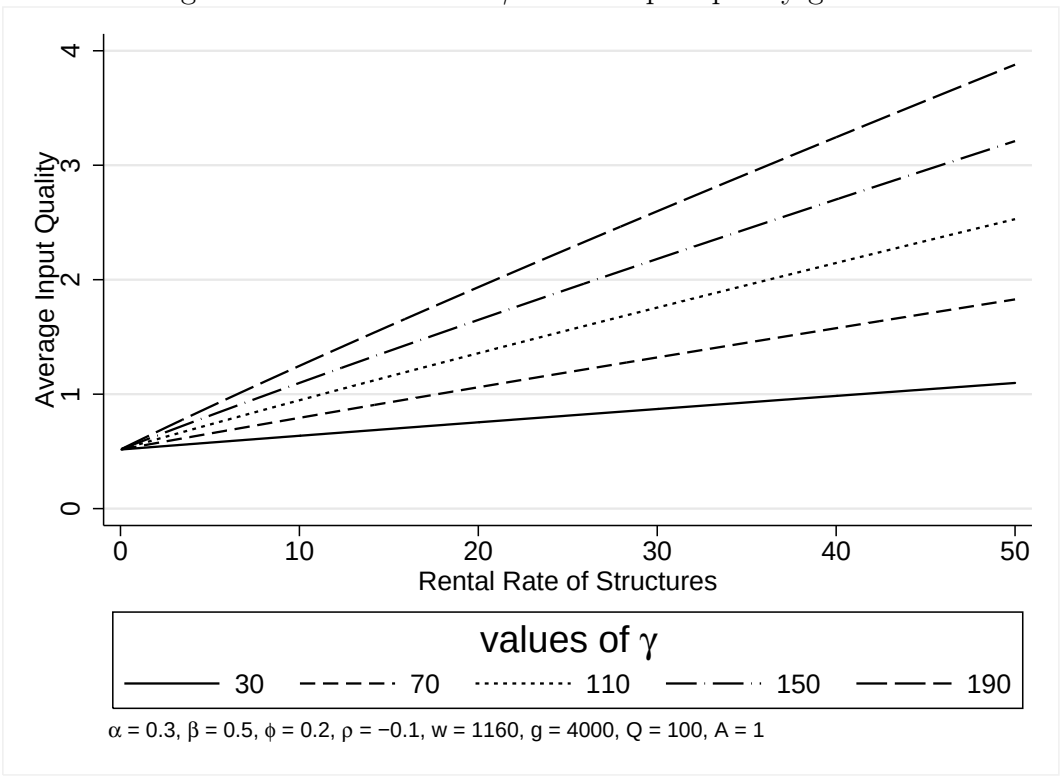


Figure 3.1: Density of firm size and employer size in the Cleveland CBSA

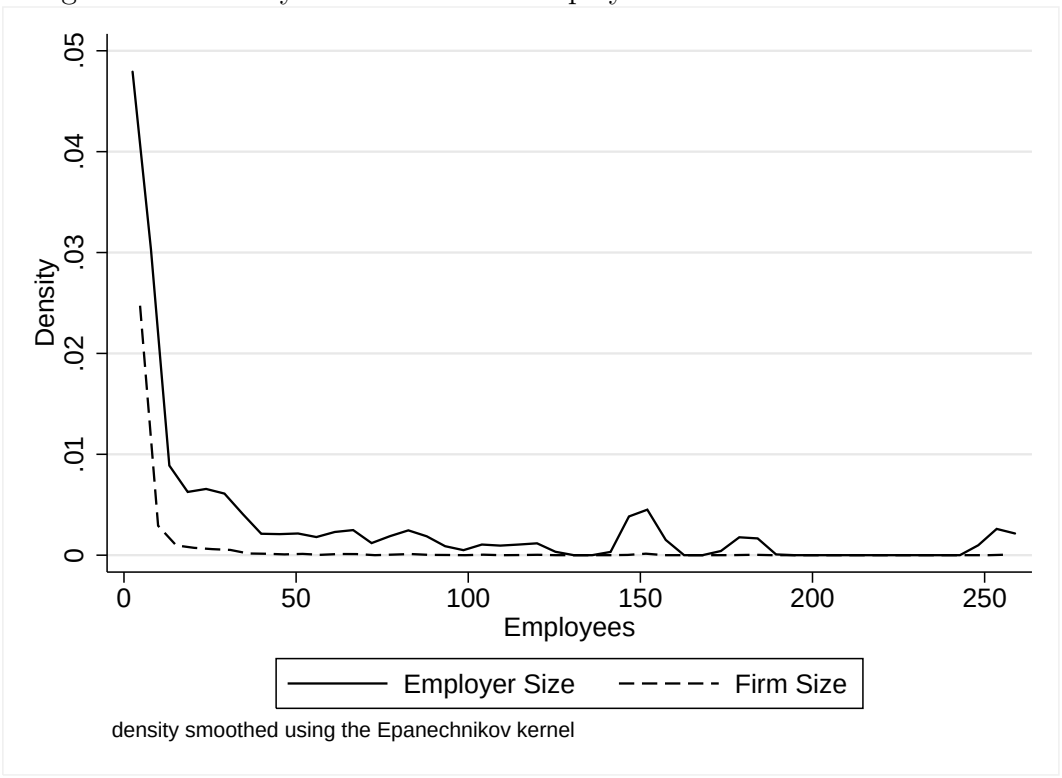


Figure 3.2: Location of private sector attorneys in the Cleveland CBSA

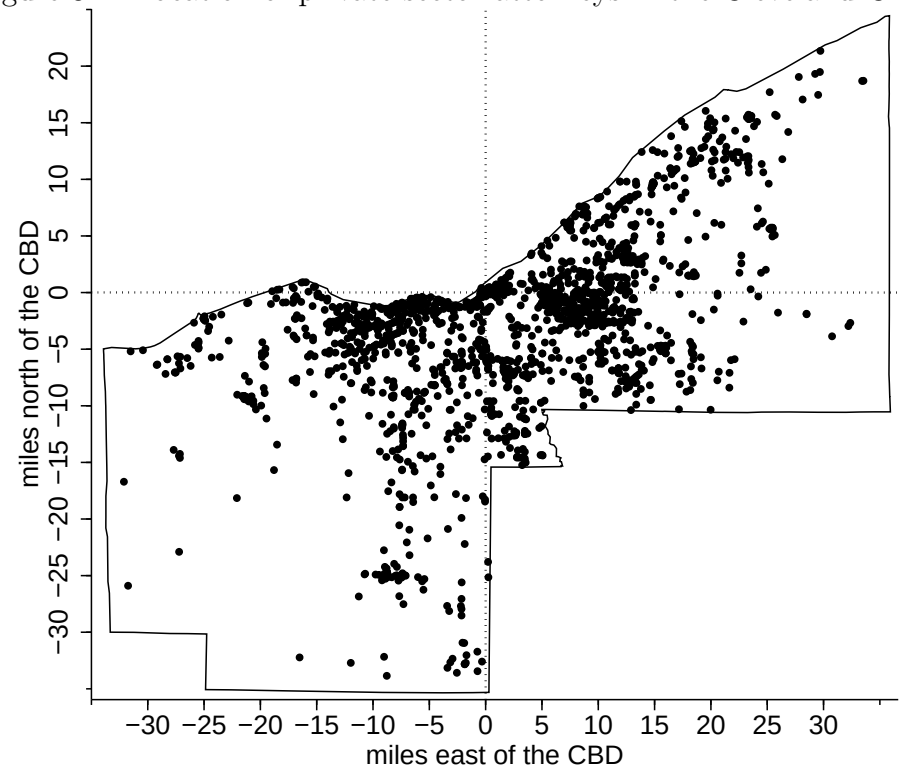


Figure 3.3: Location of CoStar properties in the Cleveland CBSA

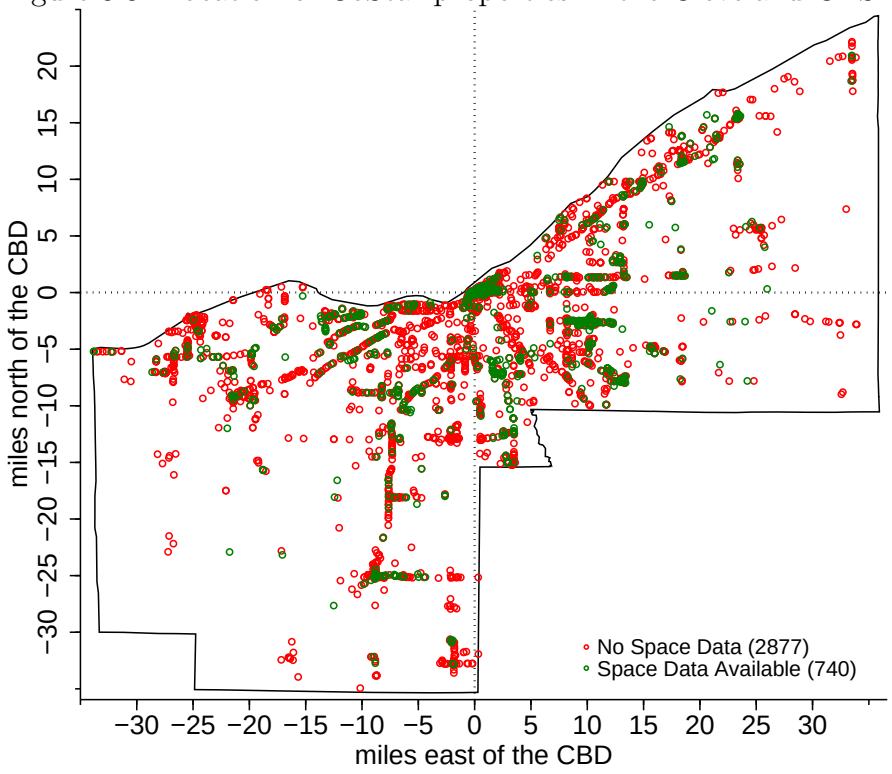


Figure 3.4: Density of Cleveland asking rents in CoStar space data

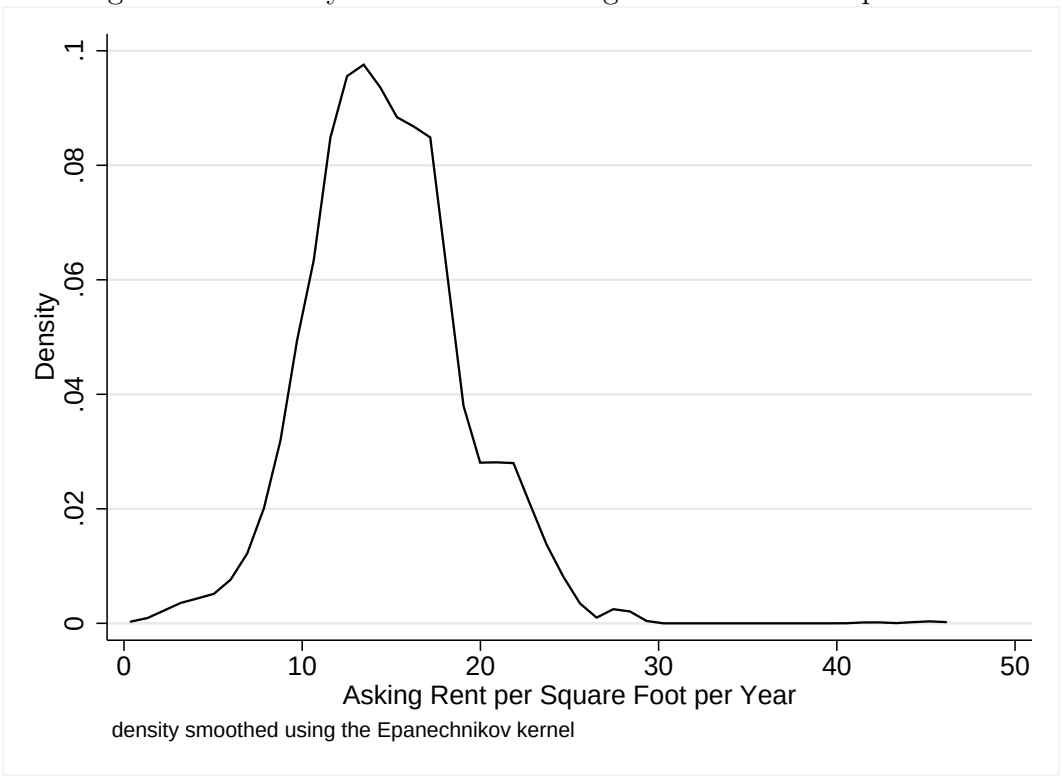
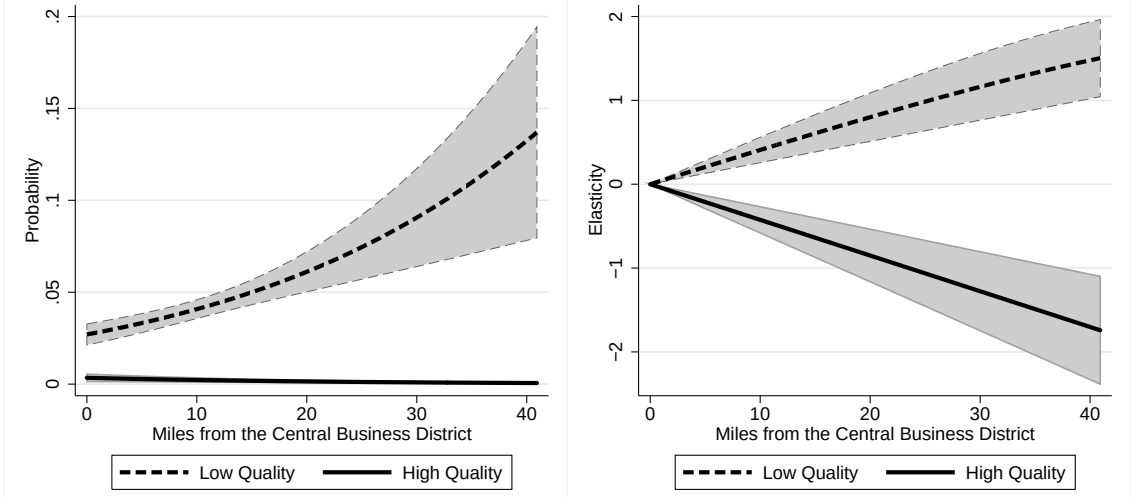
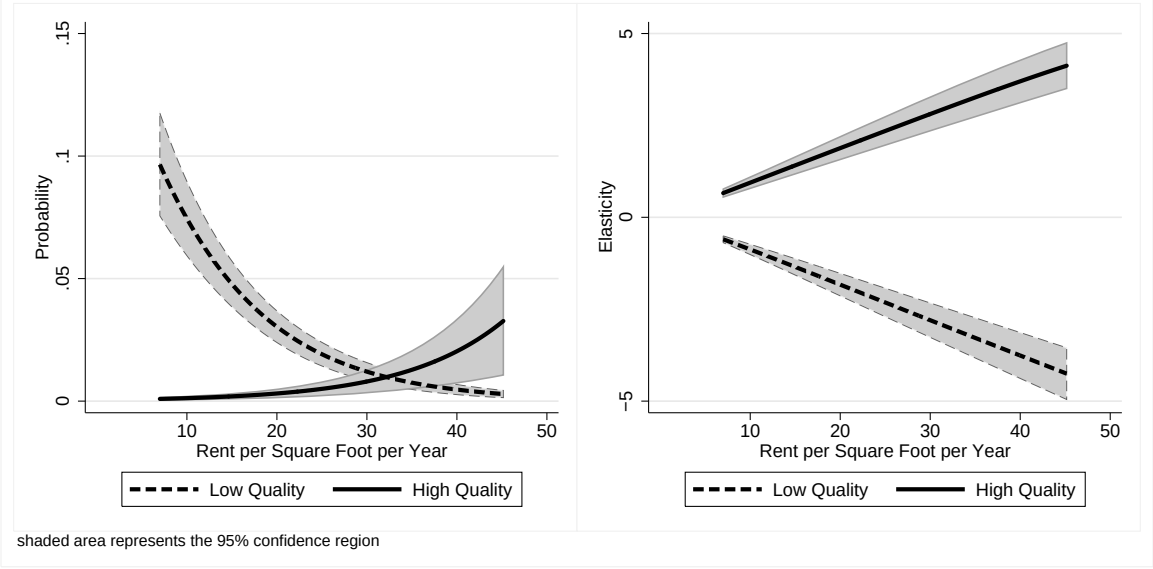


Figure 3.5: Probabilities and elasticities from the distance model



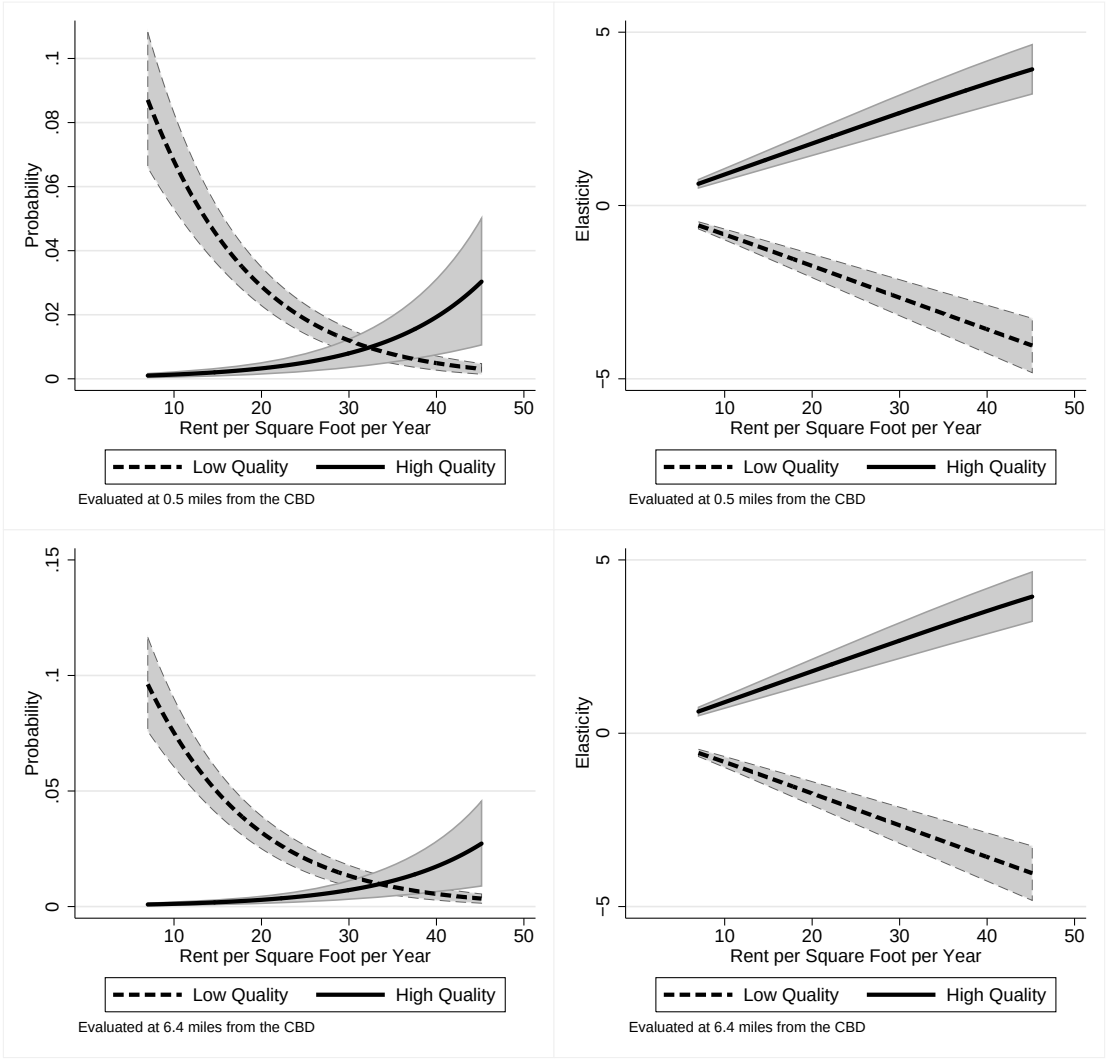
shaded area represents the 95% confidence region

Figure 3.6: Probabilities and elasticities using matched properties: model 1



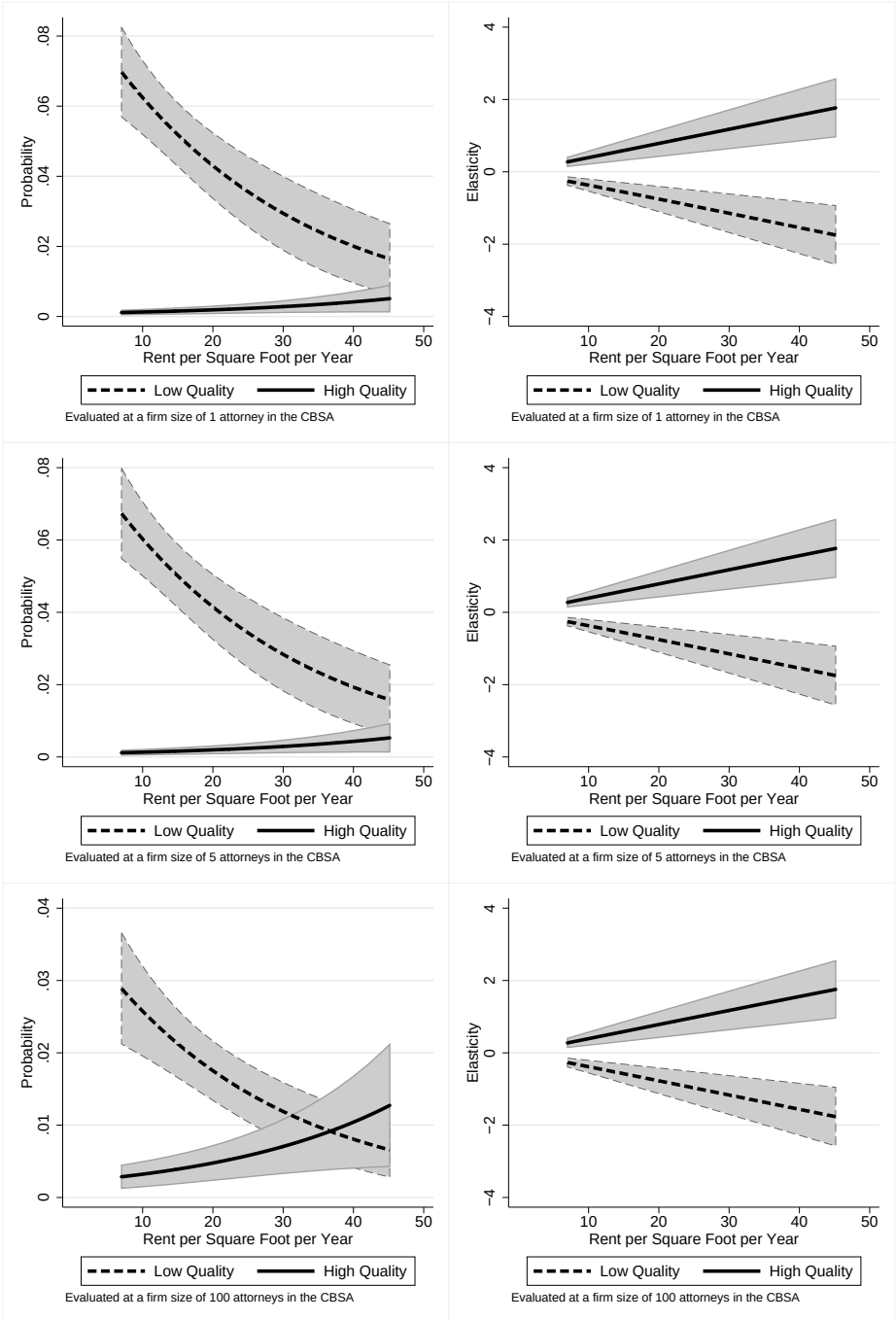
shaded area represents the 95% confidence region

Figure 3.7: Probabilities and elasticities using matched properties: model 2



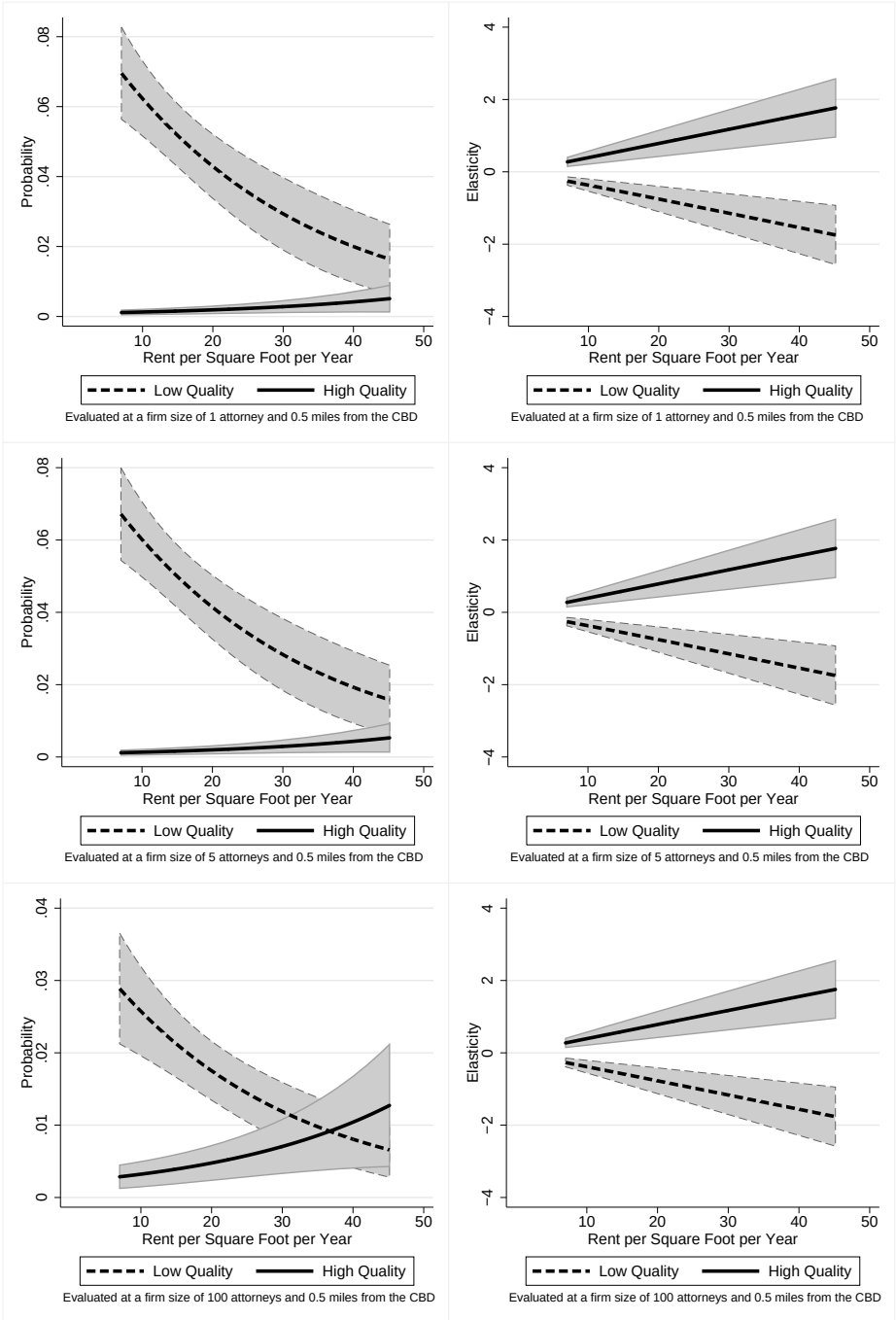
shaded area represents the 95% confidence region

Figure 3.8: Probabilities and elasticities using matched properties: model 3



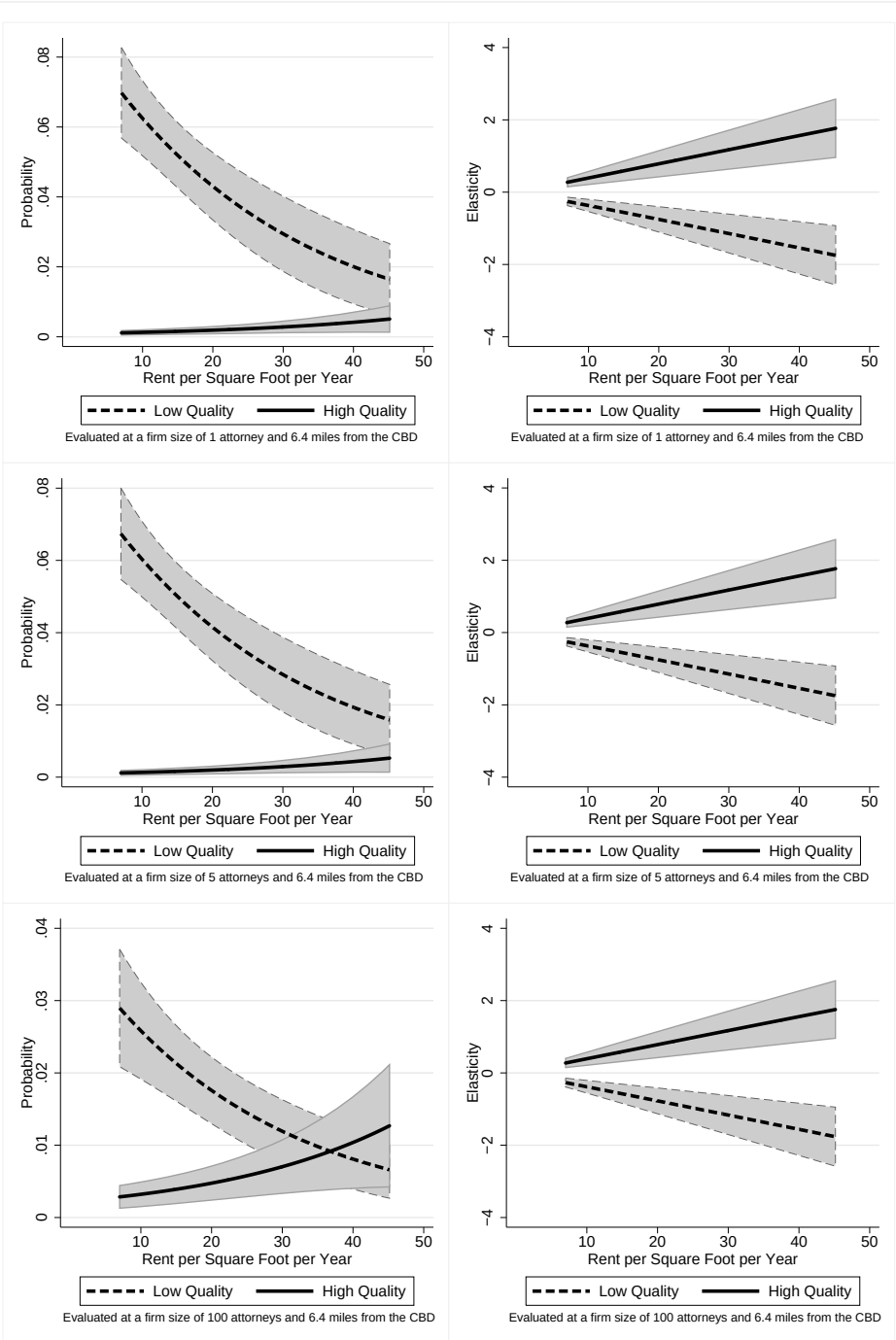
shaded area represents the 95% confidence region

Figure 3.9: Probabilities and elasticities using matched properties: model 4



shaded area represents the 95% confidence region

Figure 3.10: Probabilities and elasticities using matched properties: model 4



shaded area represents the 95% confidence region

Figure 3.11: Probabilities and elasticities using matched properties

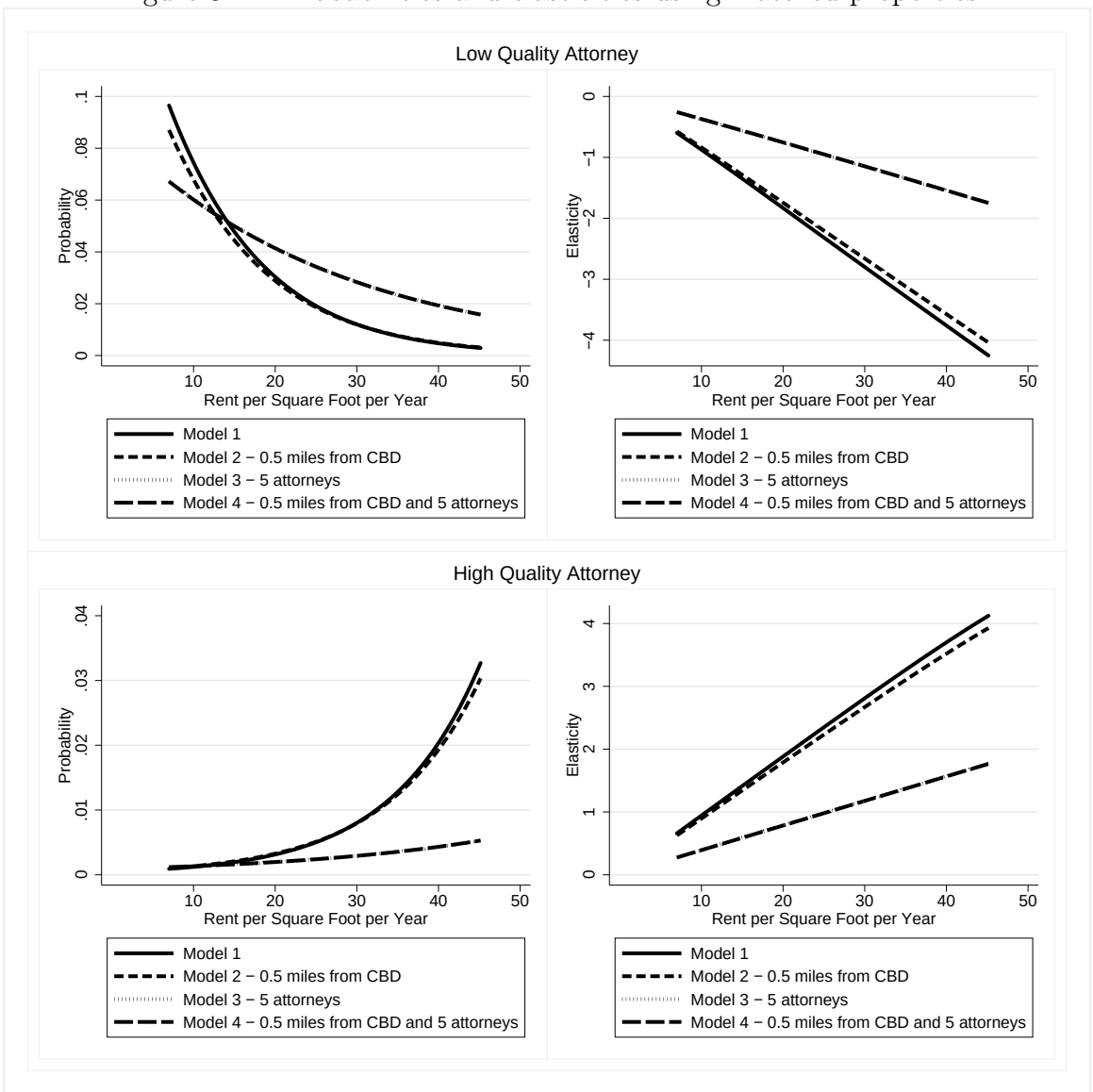


Figure 3.12: predicted rent in the Cleveland CBSA

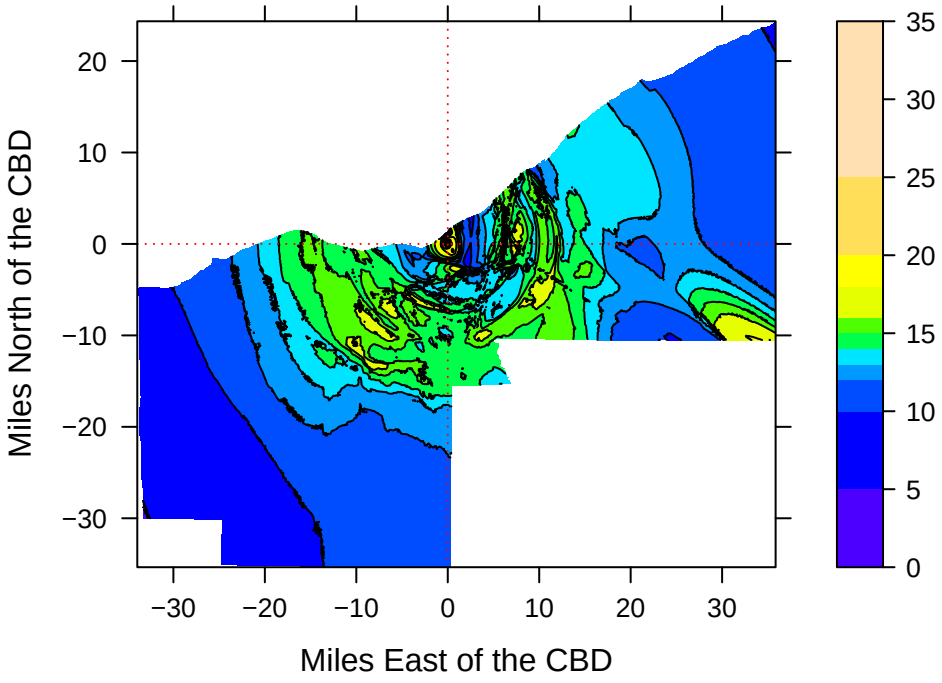


Figure 3.13: predicted rent within five miles of the Cleveland CBD

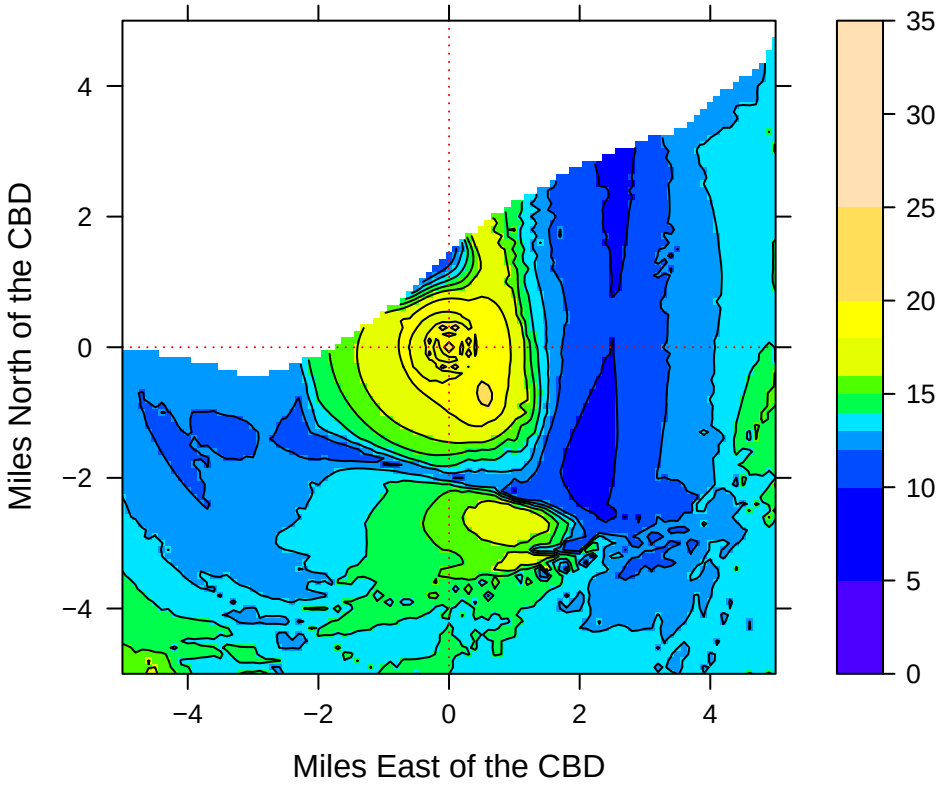
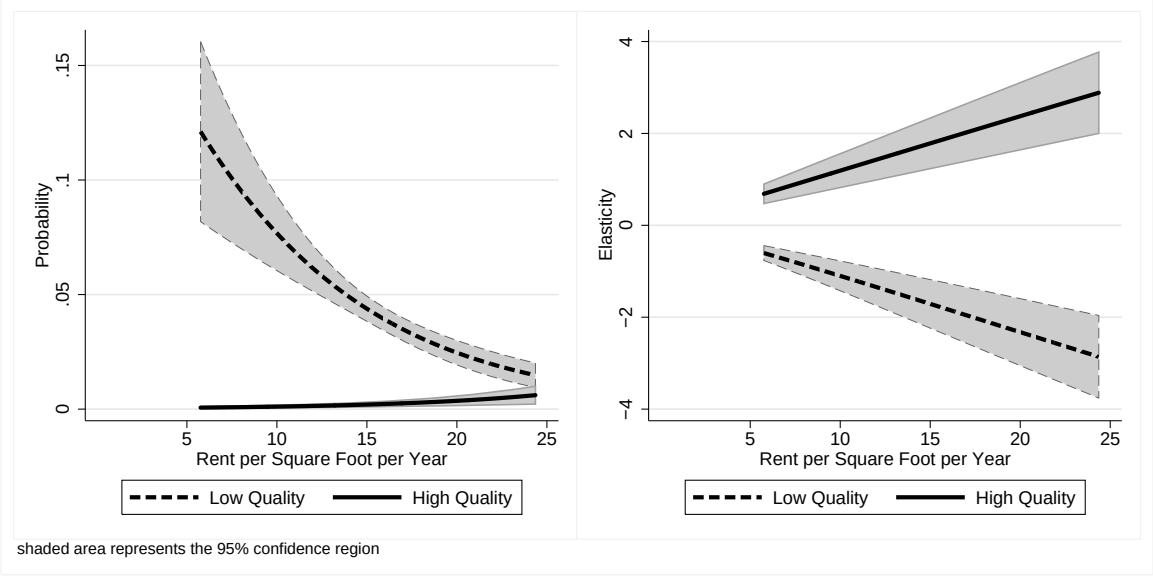
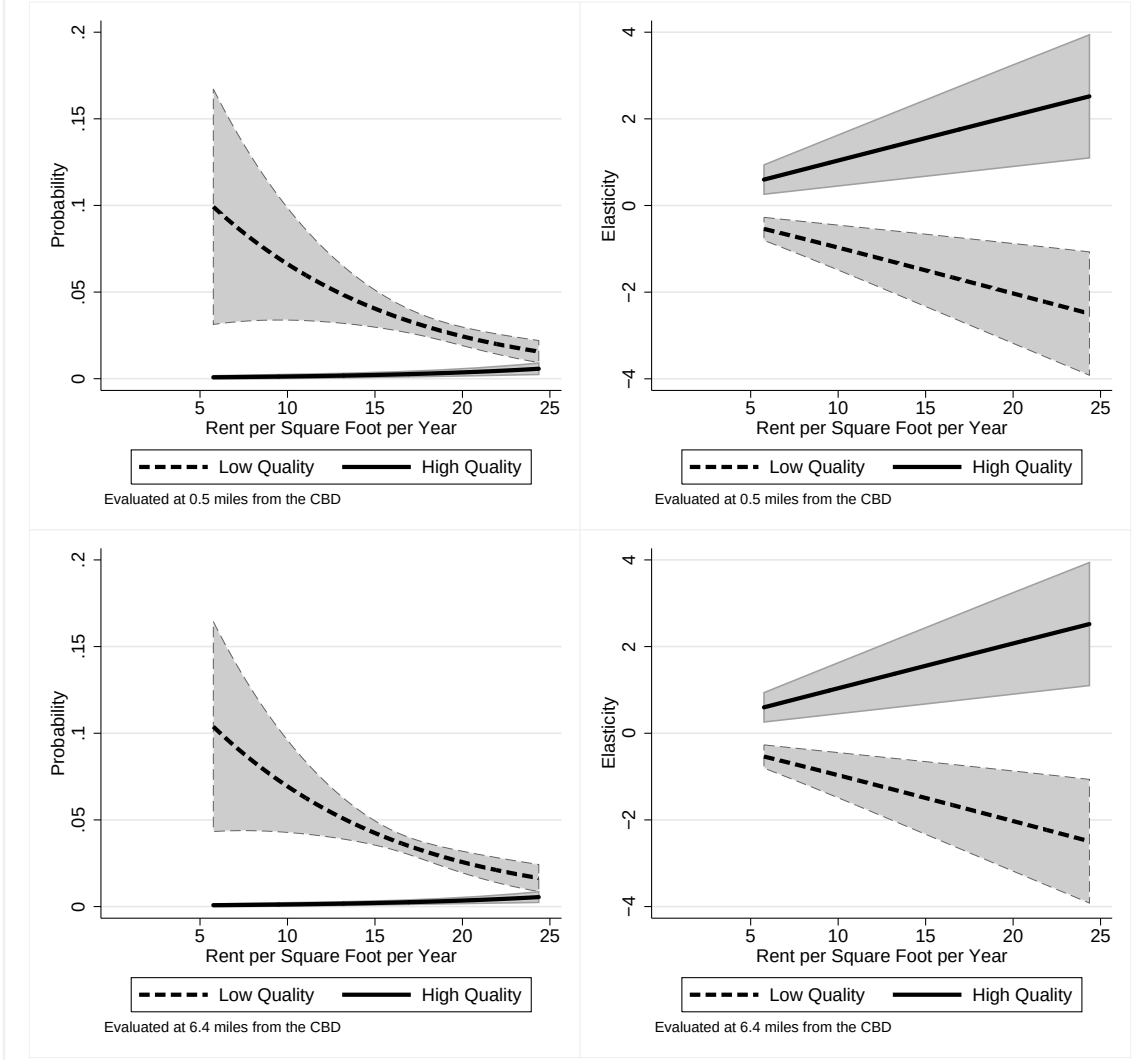


Figure 3.14: Probabilities and elasticities using predicted rent: model 1



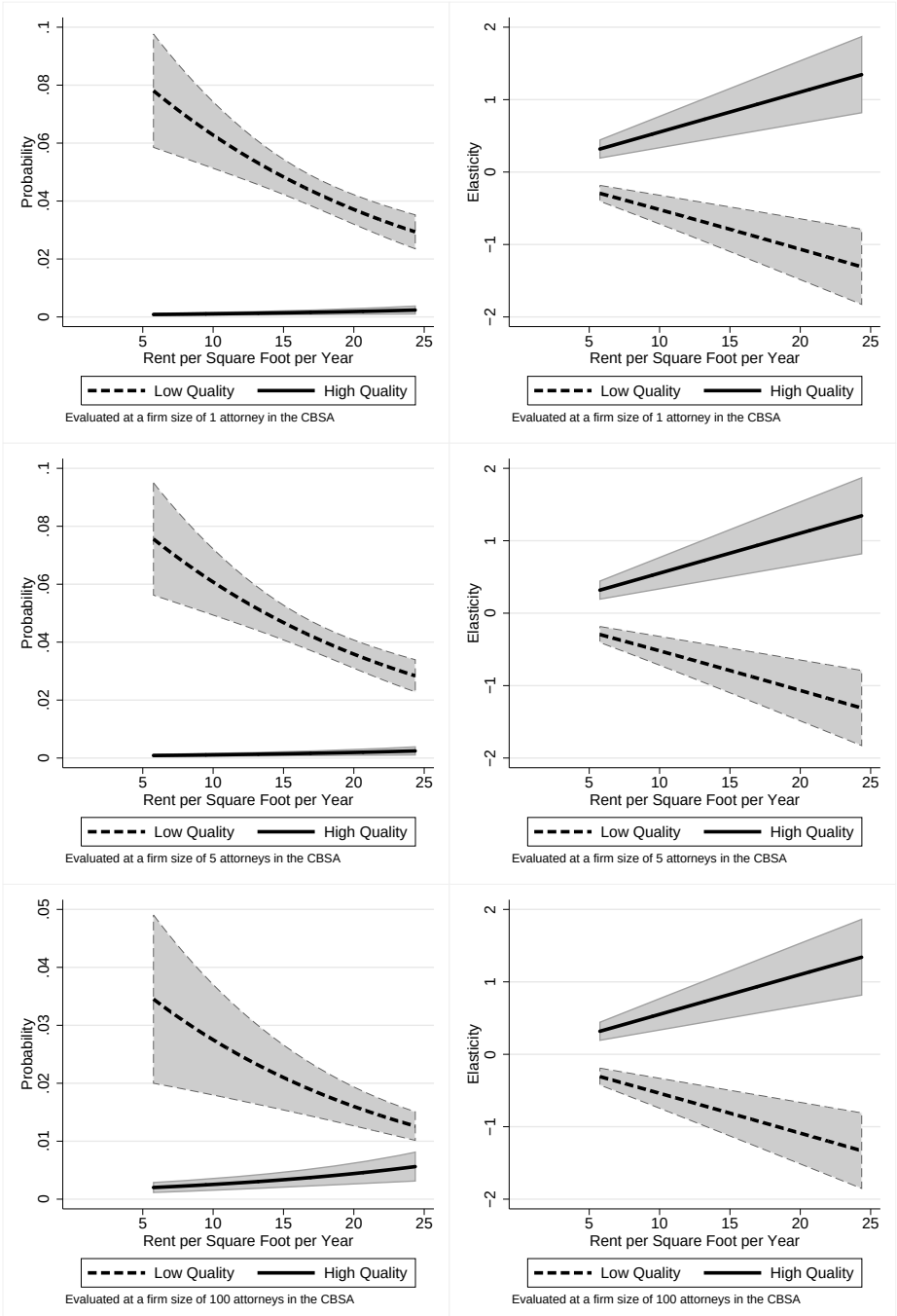
shaded area represents the 95% confidence region

Figure 3.15: Probabilities and elasticities using predicted rent: model 2



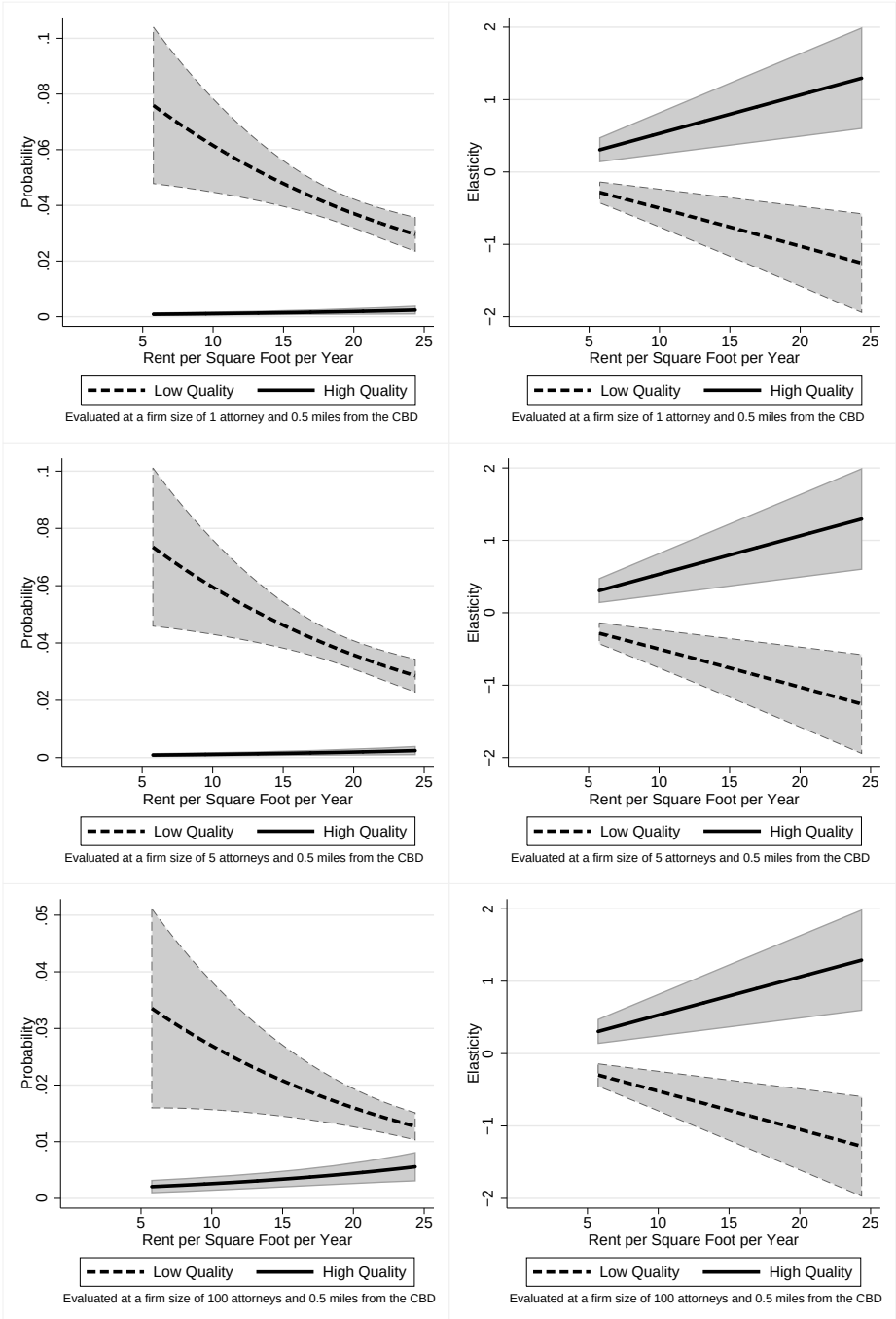
shaded area represents the 95% confidence region

Figure 3.16: Probabilities and elasticities using predicted rent: model 3



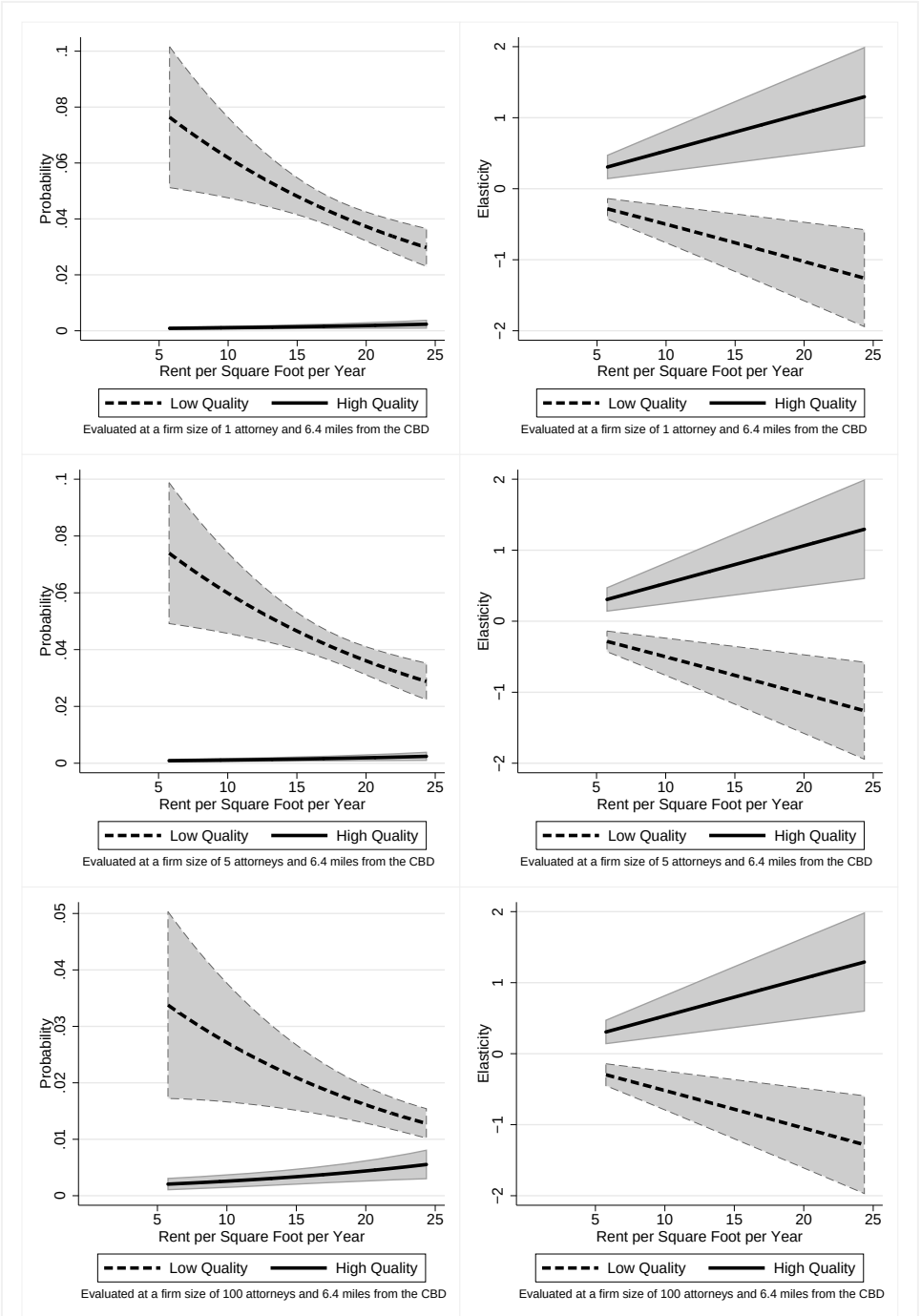
shaded area represents the 95% confidence region

Figure 3.17: Probabilities and elasticities using predicted rent: model 4



shaded area represents the 95% confidence region

Figure 3.18: Probabilities and elasticities using predicted rent: model 4



shaded area represents the 95% confidence region

Figure 3.19: Probabilities and elasticities using predicted rent

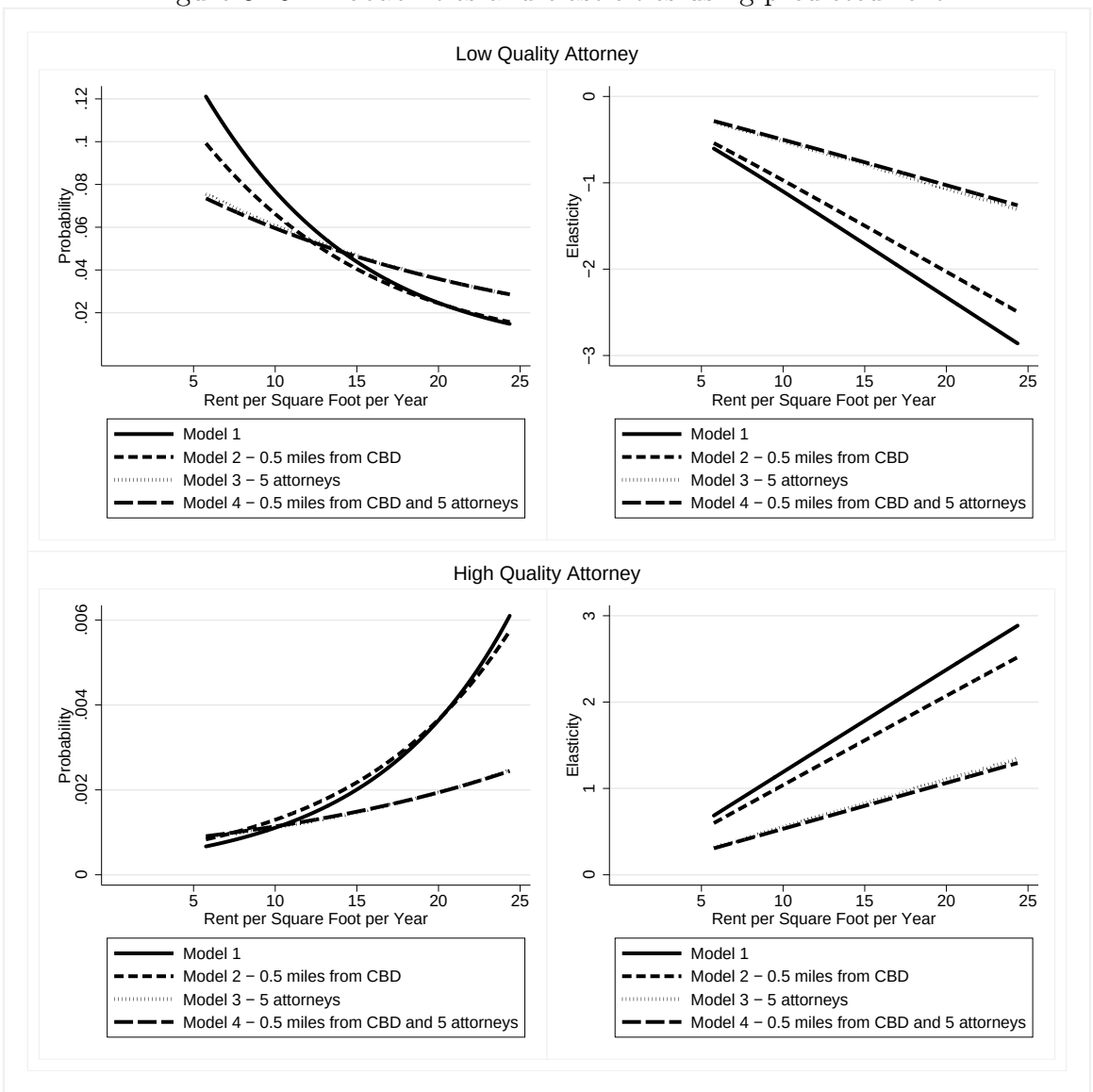
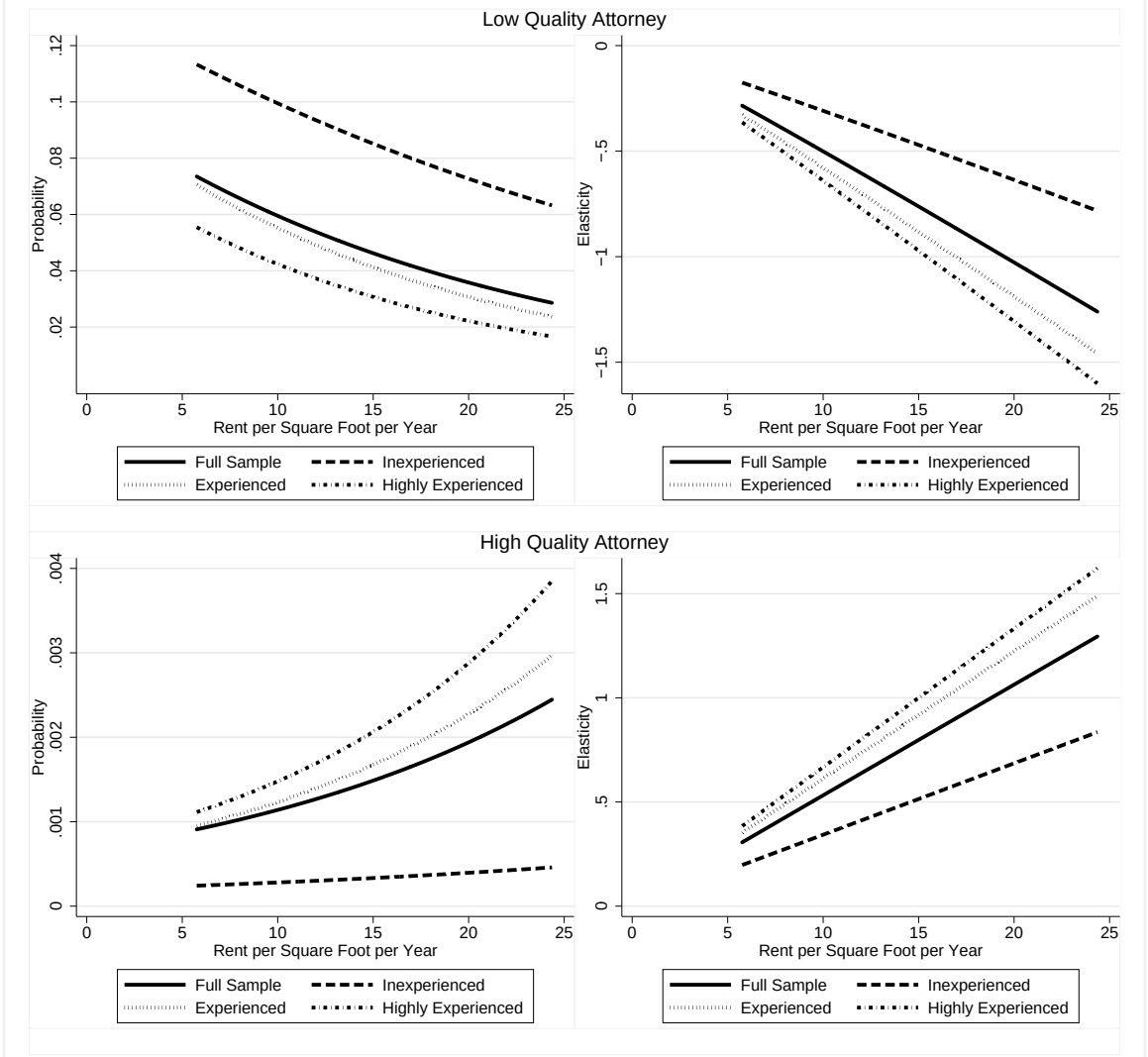


Figure 3.20: Probabilities and elasticities using predicted rent: effect of experience



Evaluated for firms employing 5 attorneys and located 0.5 miles from CBD

Figure 3.21: Elasticities using predicted rent: effect of school location

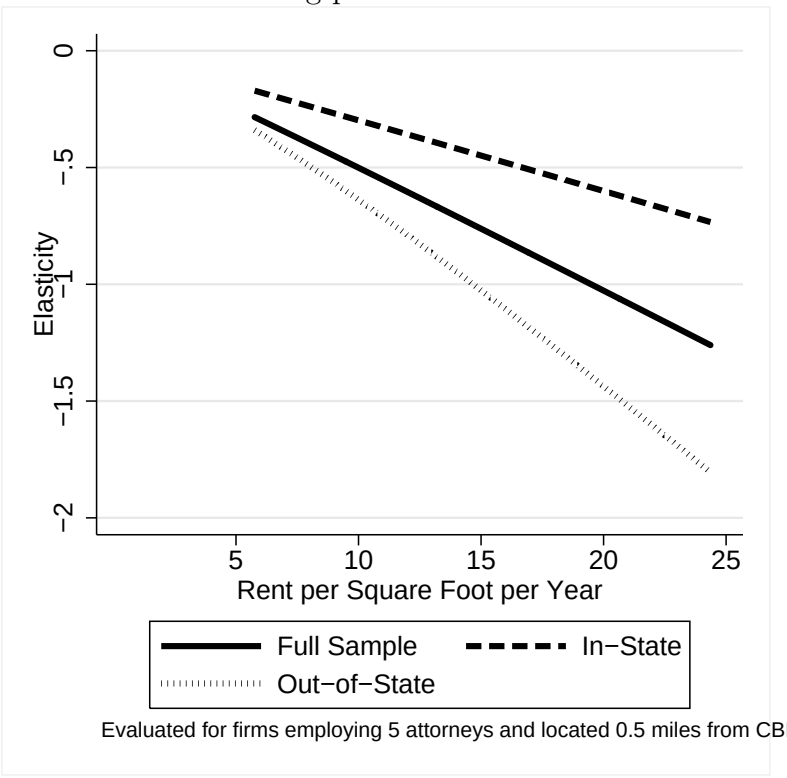


Figure 4.1: Estimated wage gradients in the Cleveland CBSA

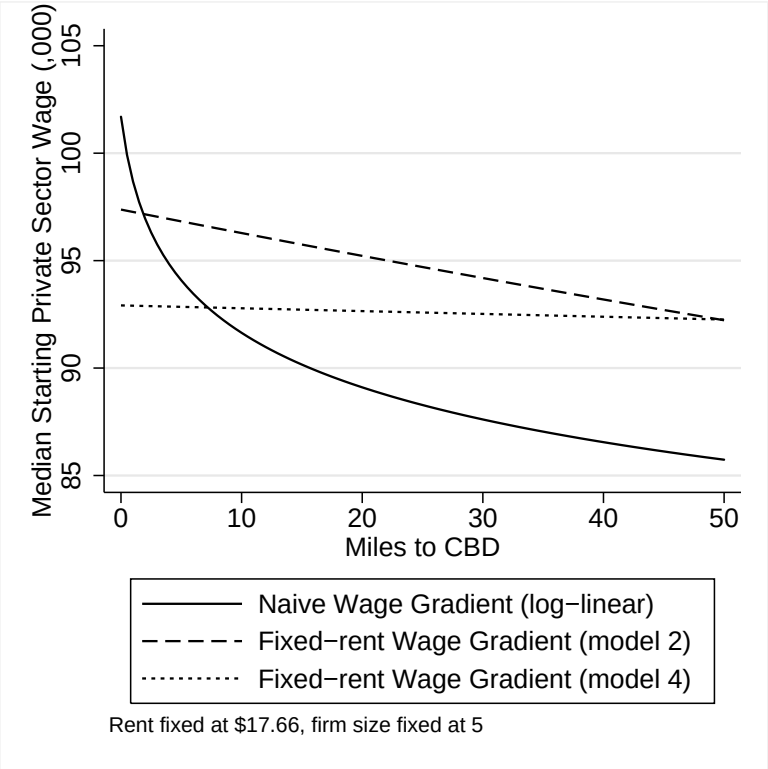


Figure B.1: Probabilities and elasticities using predicted rent and alternative error structures: model 1

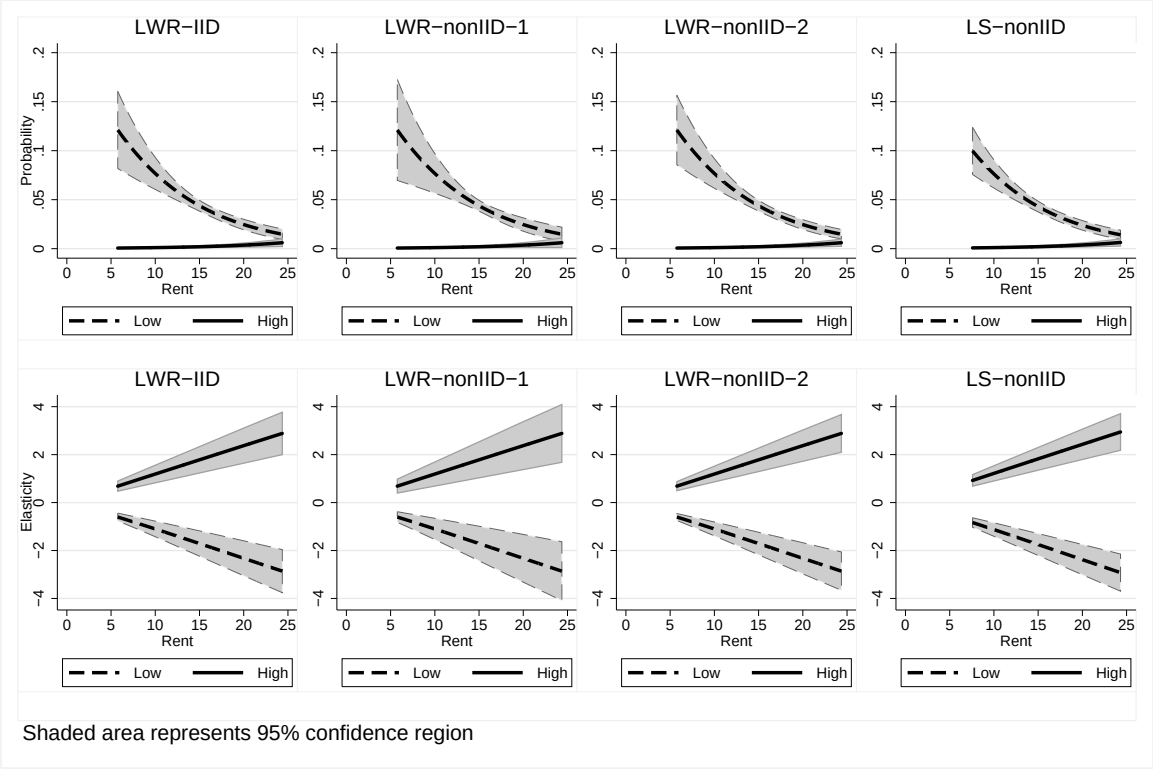


Figure B.2: Probabilities and elasticities using predicted rent and alternative error structures: model 2

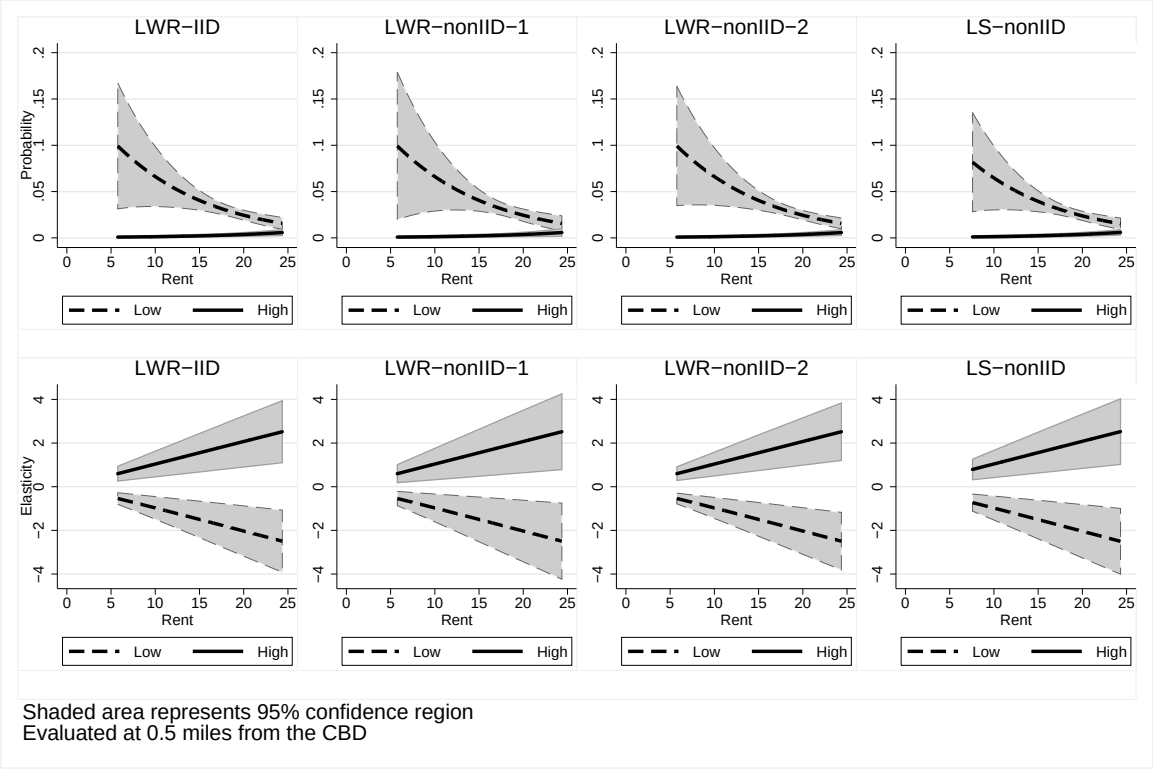


Figure B.3: Probabilities and elasticities using predicted rent and alternative error structures: model 2

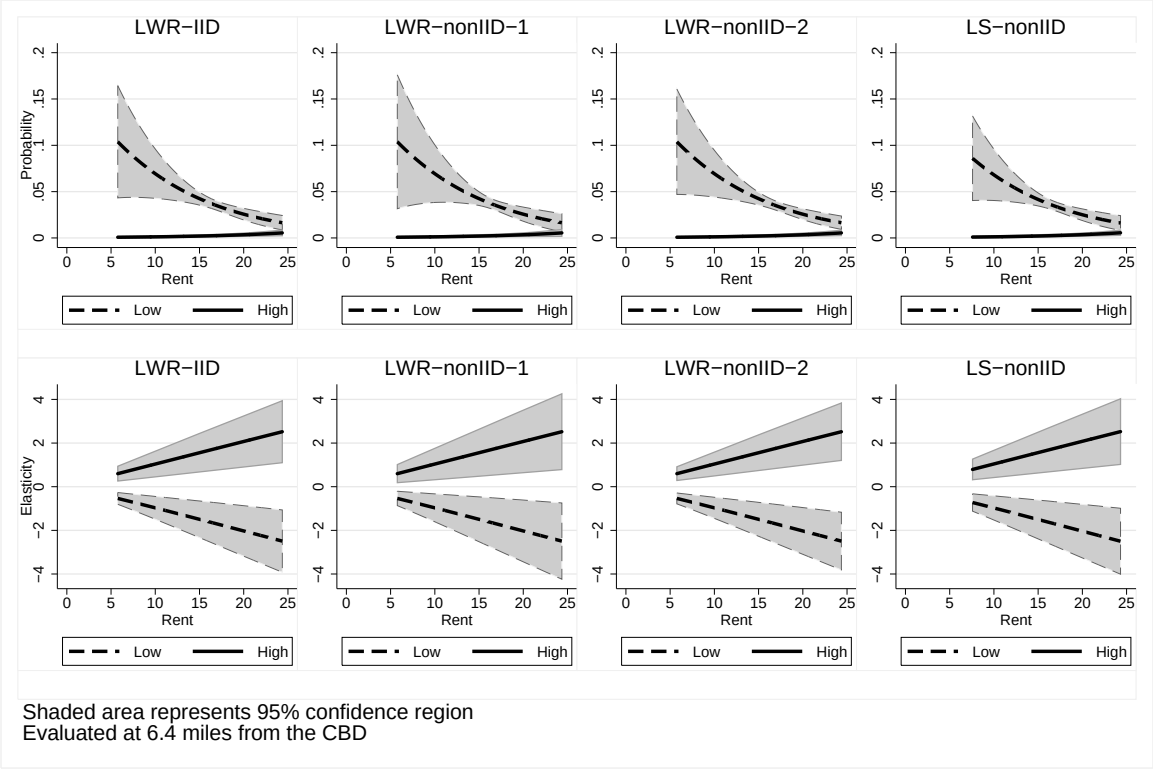


Figure B.4: Probabilities and elasticities using predicted rent and alternative error structures: model 3

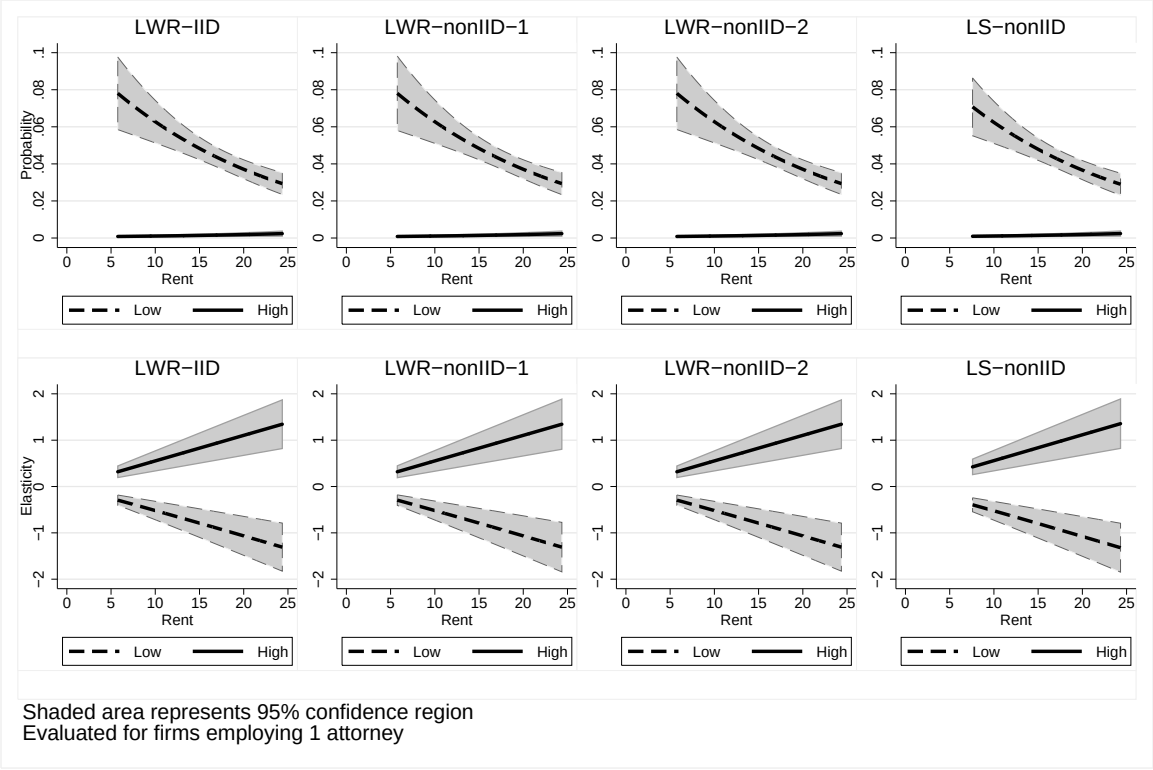


Figure B.5: Probabilities and elasticities using predicted rent and alternative error structures: model 3

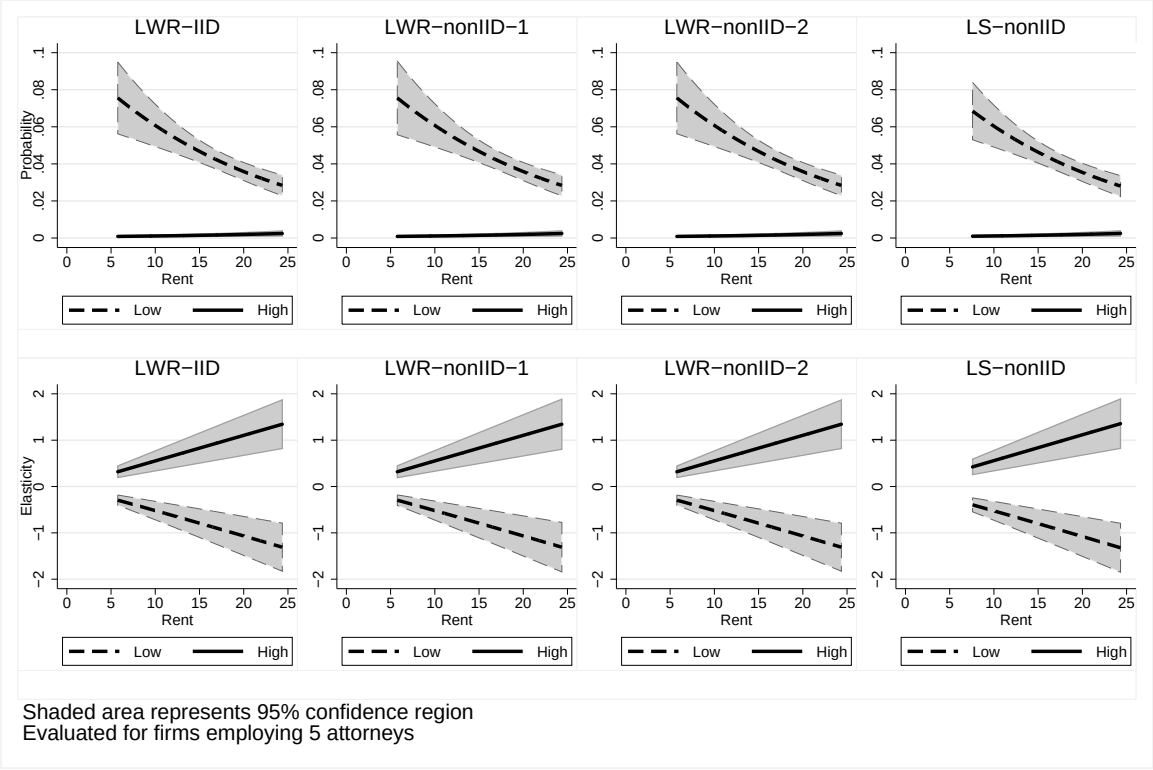


Figure B.6: Probabilities and elasticities using predicted rent and alternative error structures: model 3

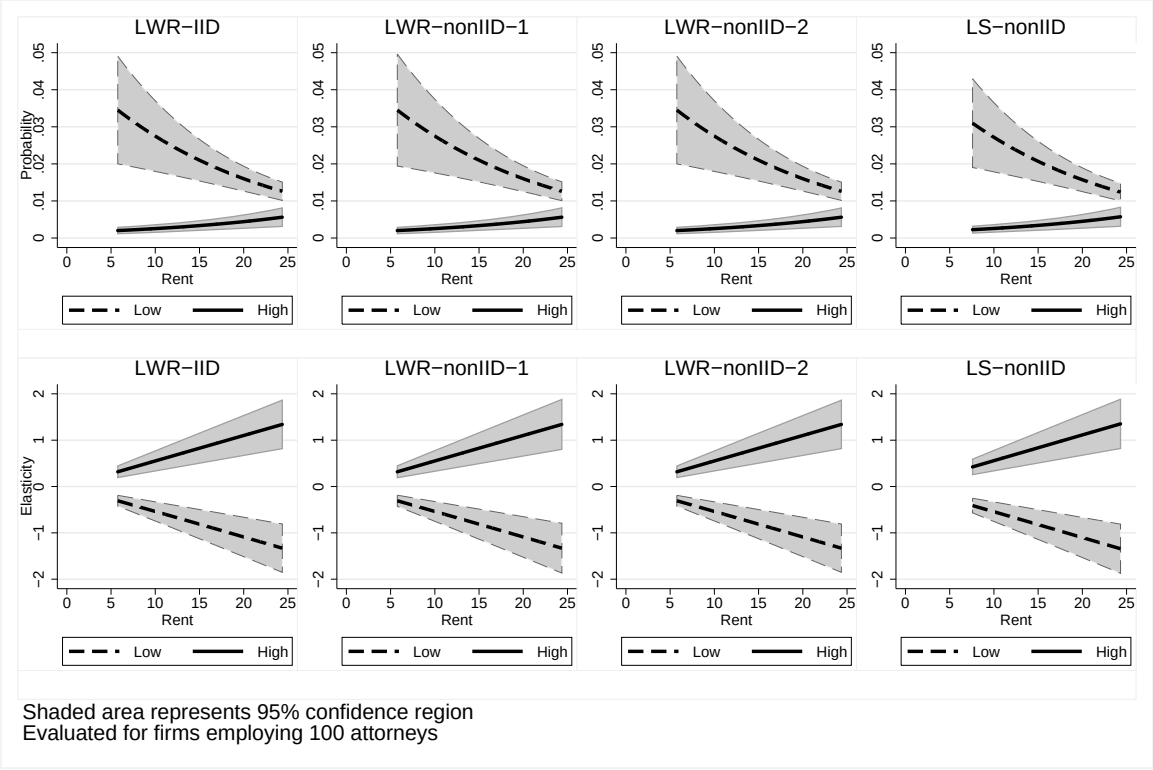


Figure B.7: Probabilities and elasticities using predicted rent and alternative error structures: model 4

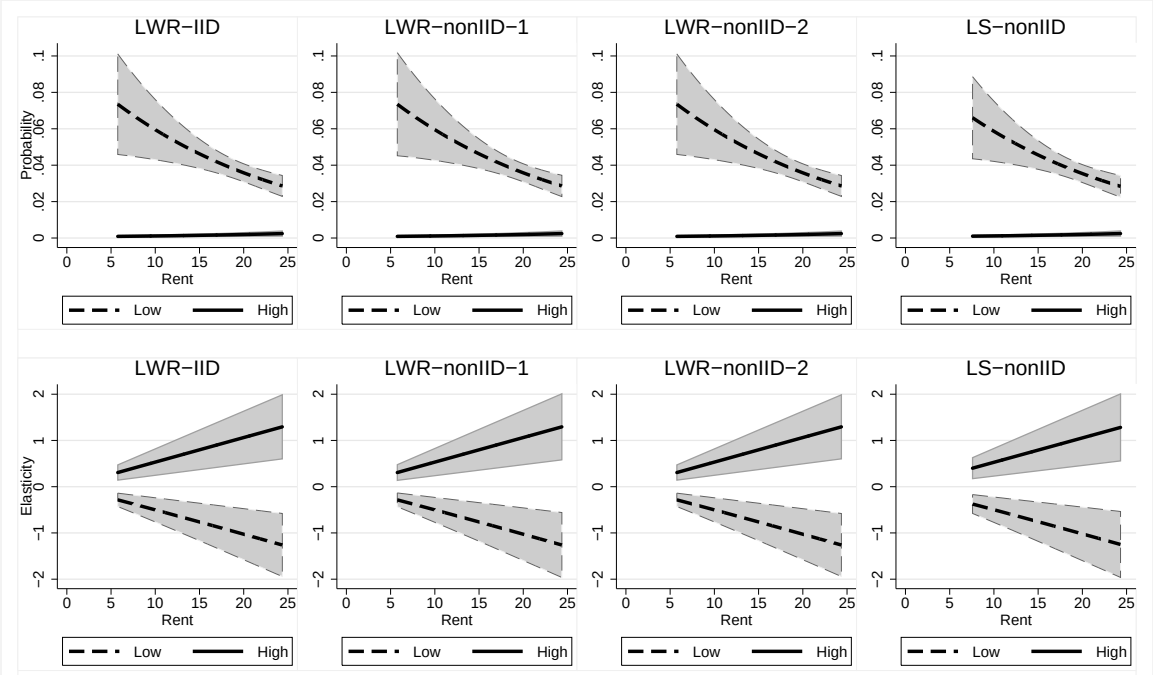


Figure B.8: Probabilities and elasticities using predicted rent: weighted and unweighted results for model 2

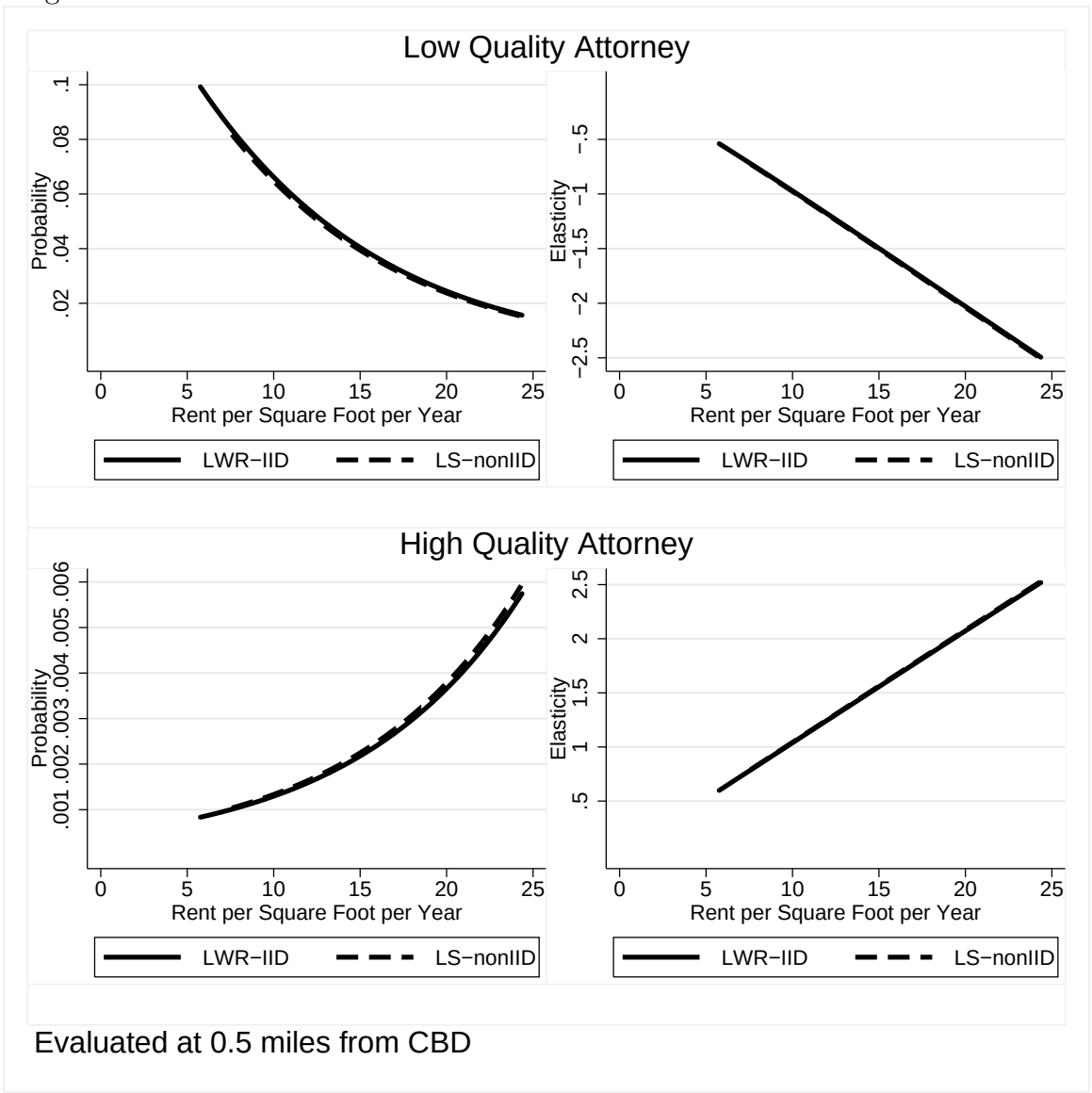


Figure B.9: Probabilities and elasticities using predicted rent: weighted and unweighted results for model 4

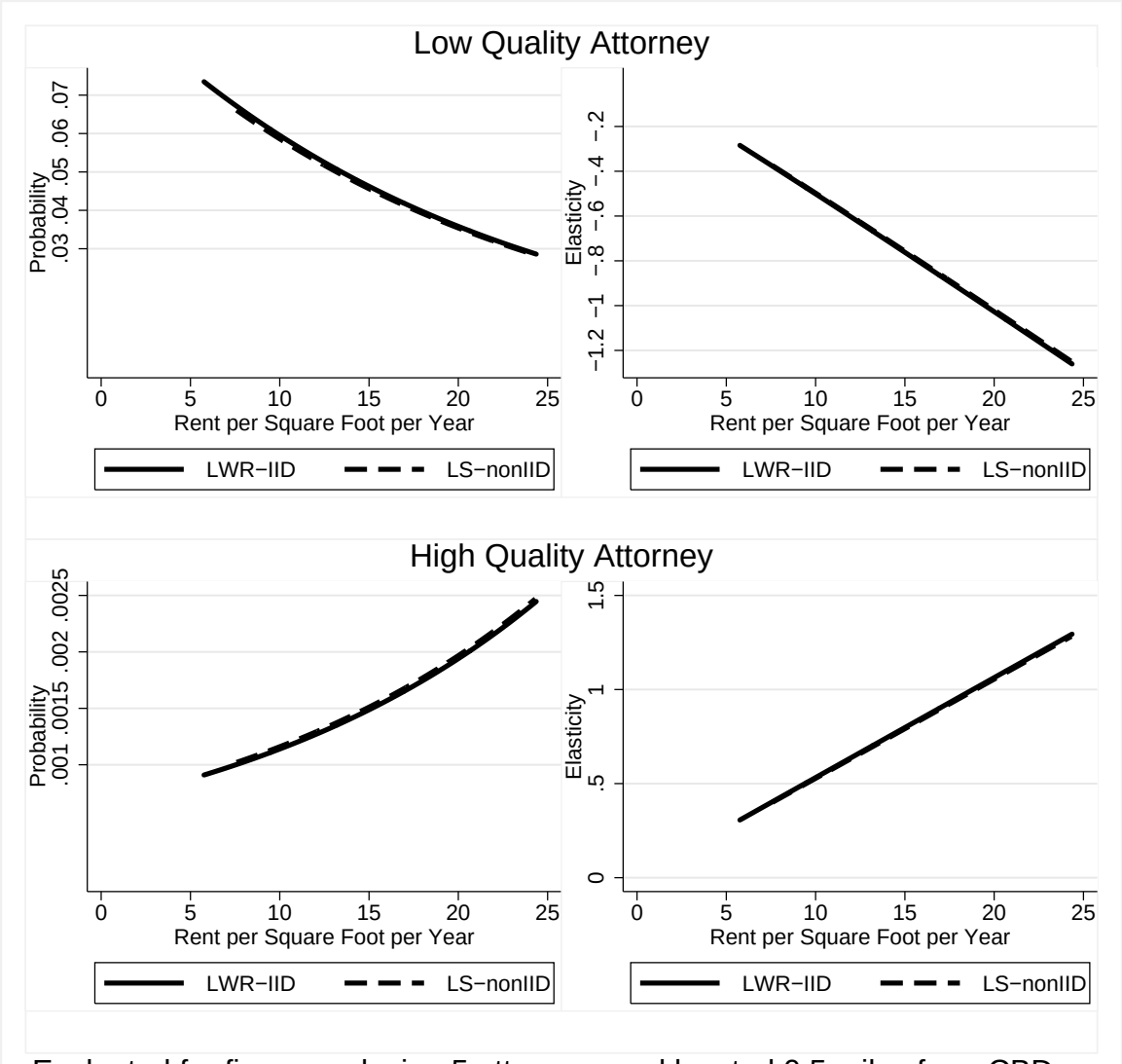


Figure C.1: Probabilities and elasticities using predicted rent and alternative bandwidths: model 1

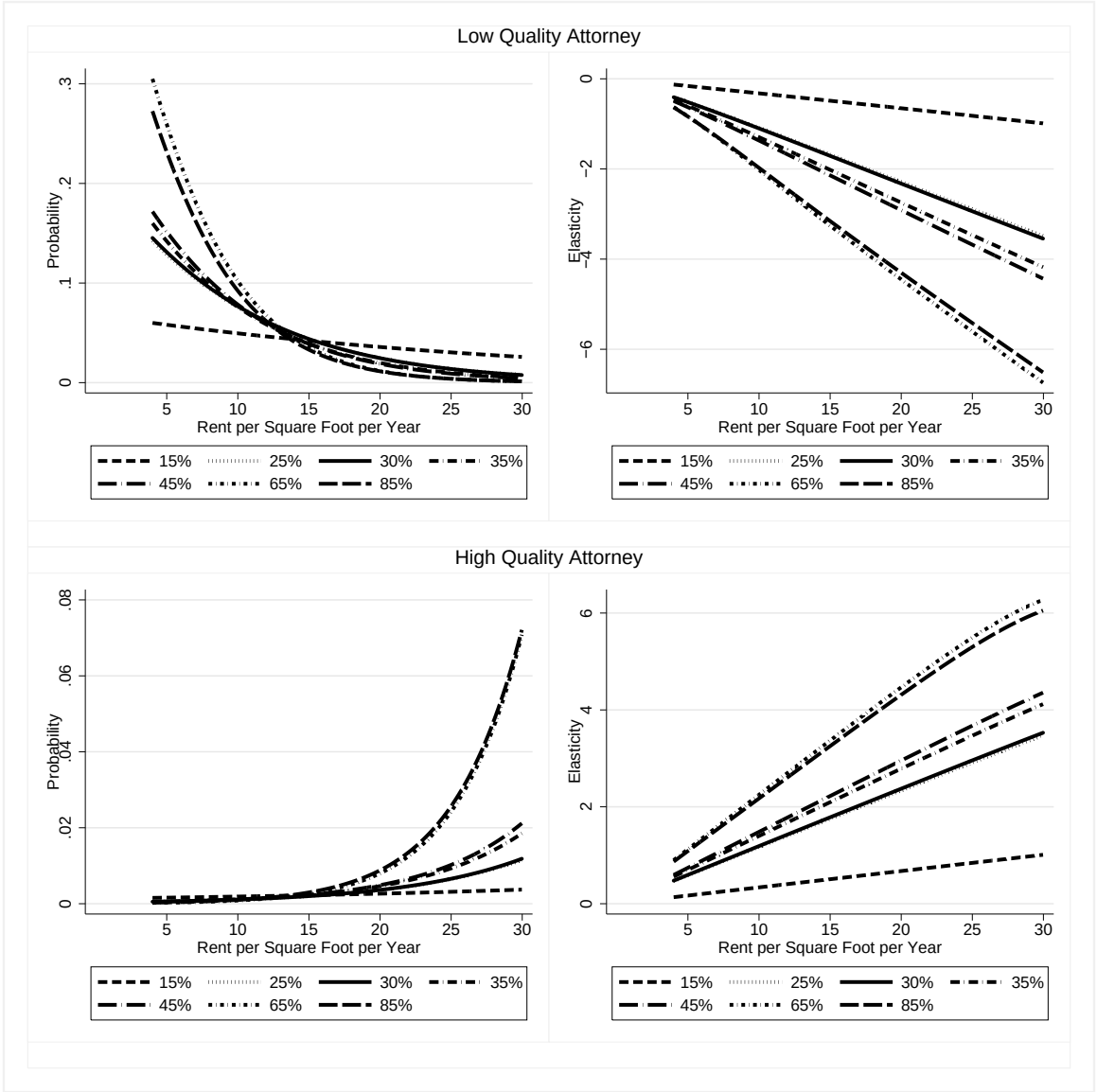
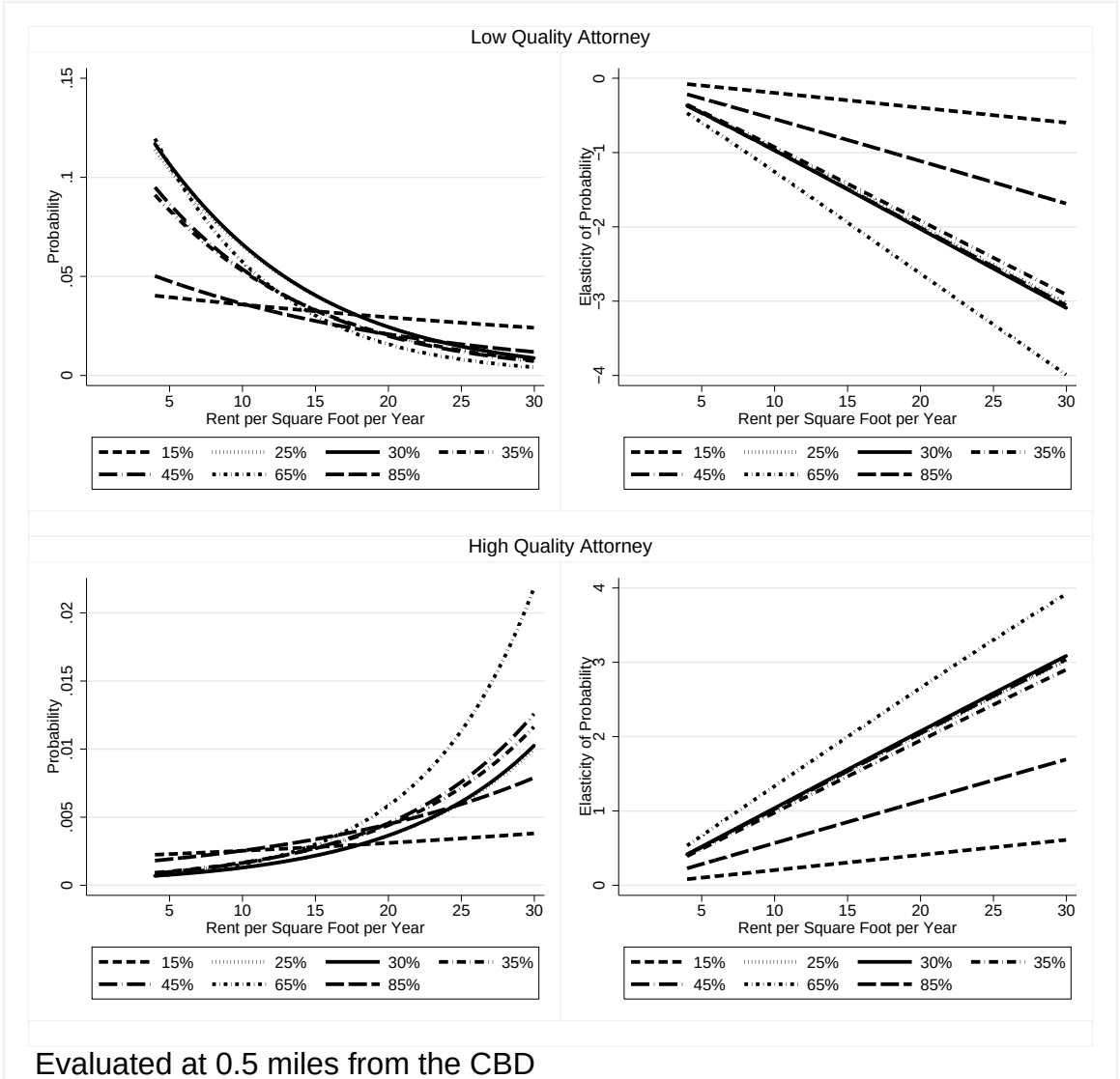


Figure C.2: Probabilities and elasticities using predicted rent and alternative bandwidths: model 2



Evaluated at 0.5 miles from the CBD

Figure C.3: Probabilities and elasticities using predicted rent and alternative bandwidths: model 2

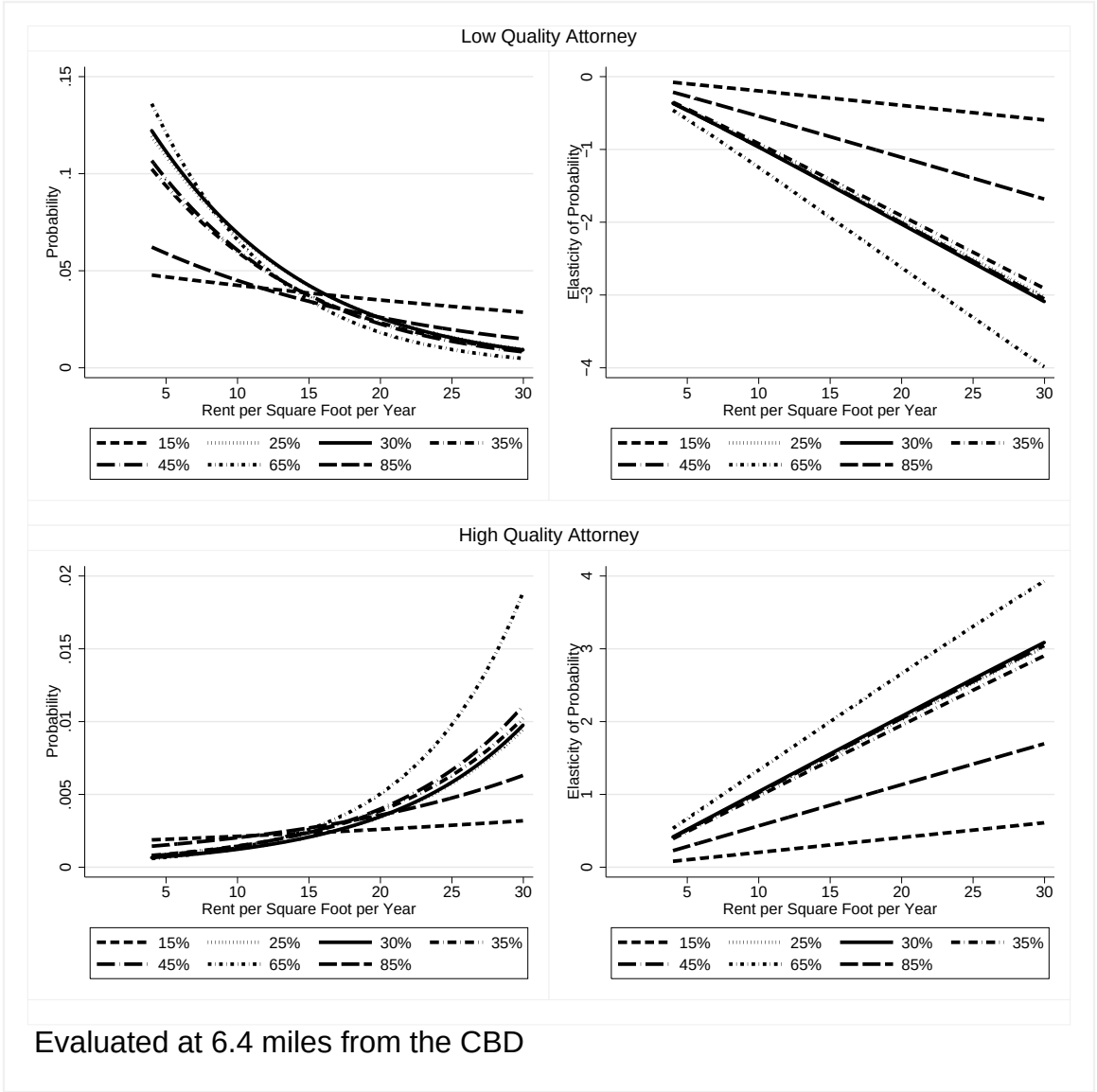
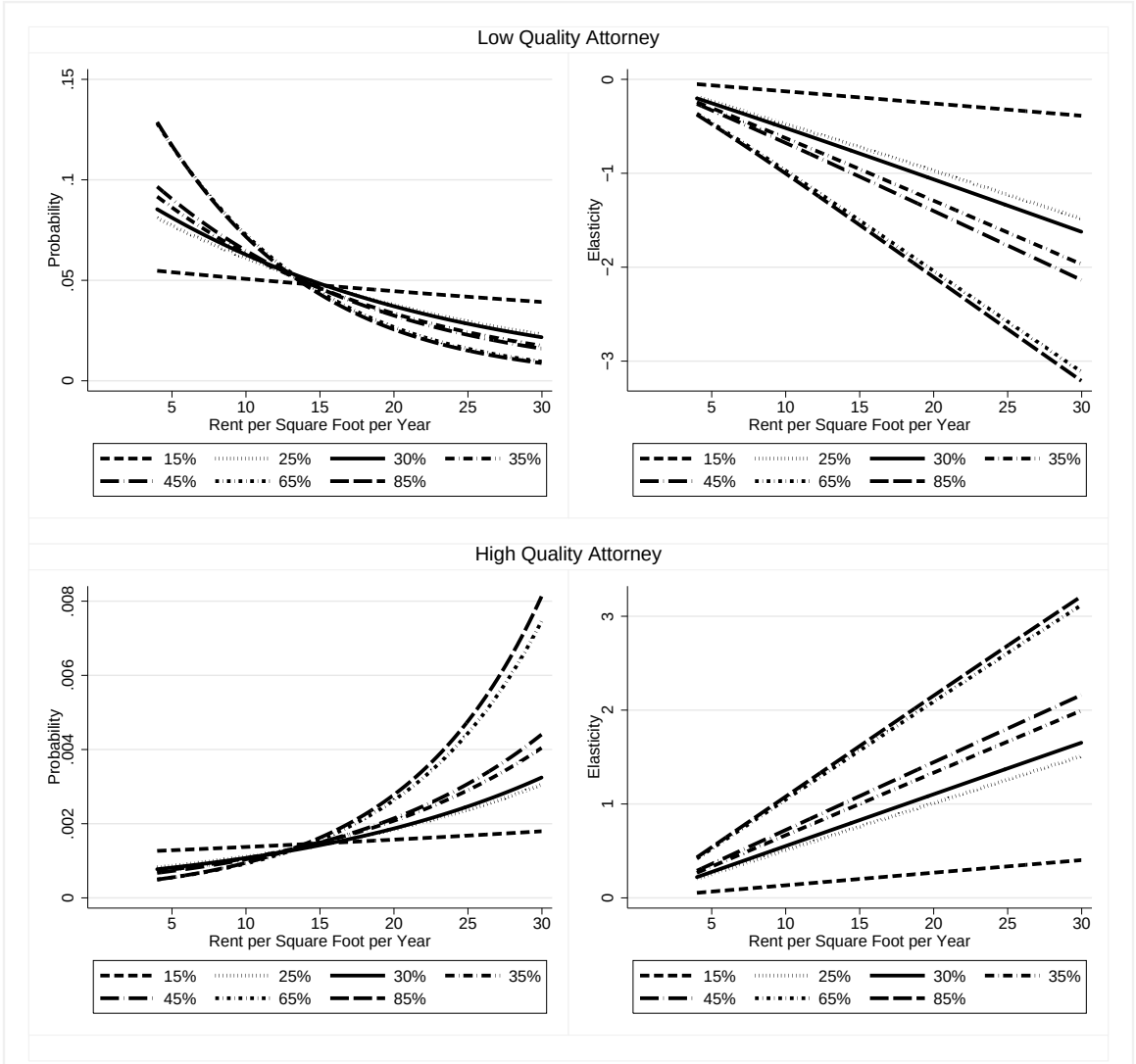
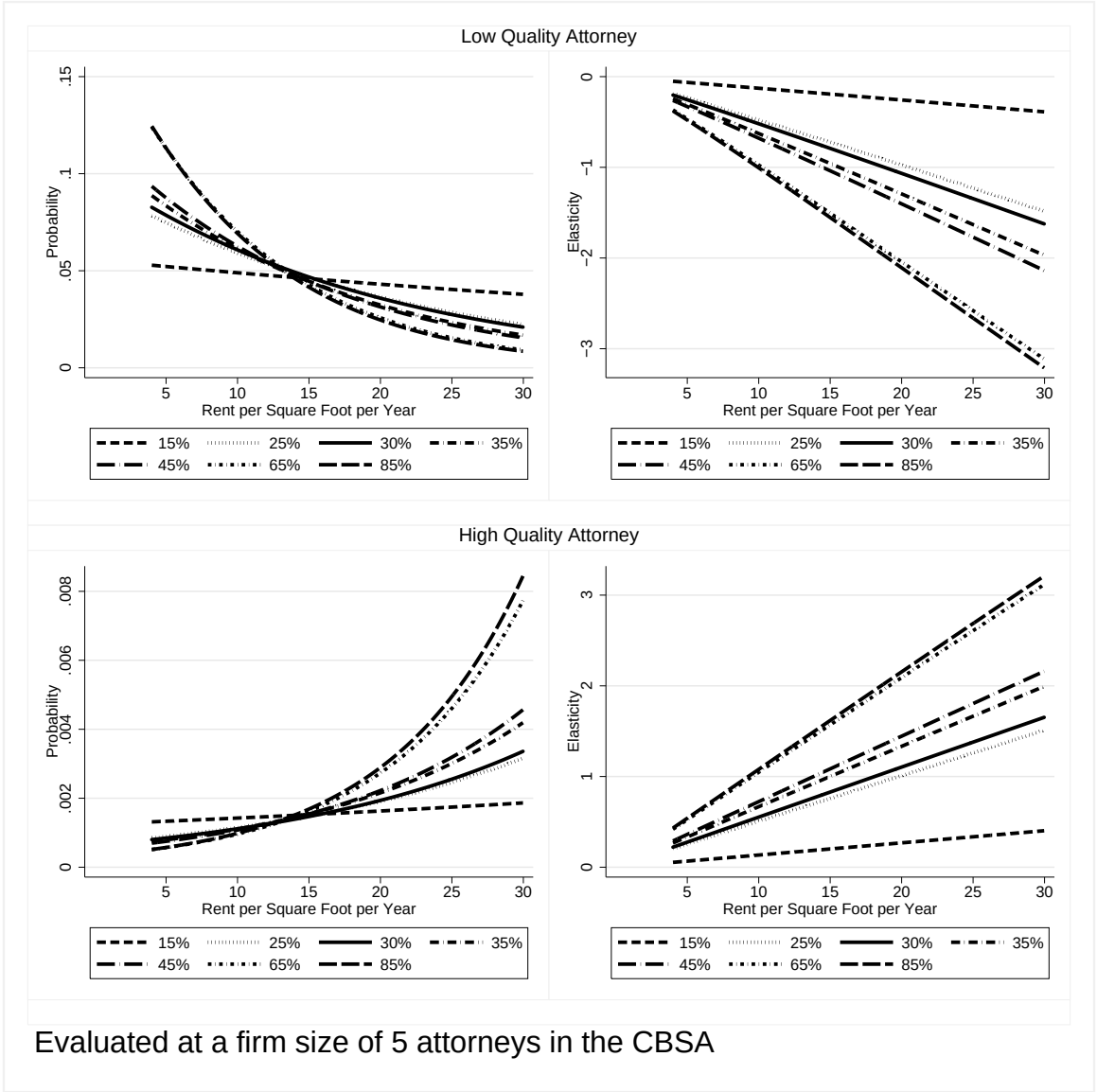


Figure C.4: Probabilities and elasticities using predicted rent and alternative bandwidths: model 3



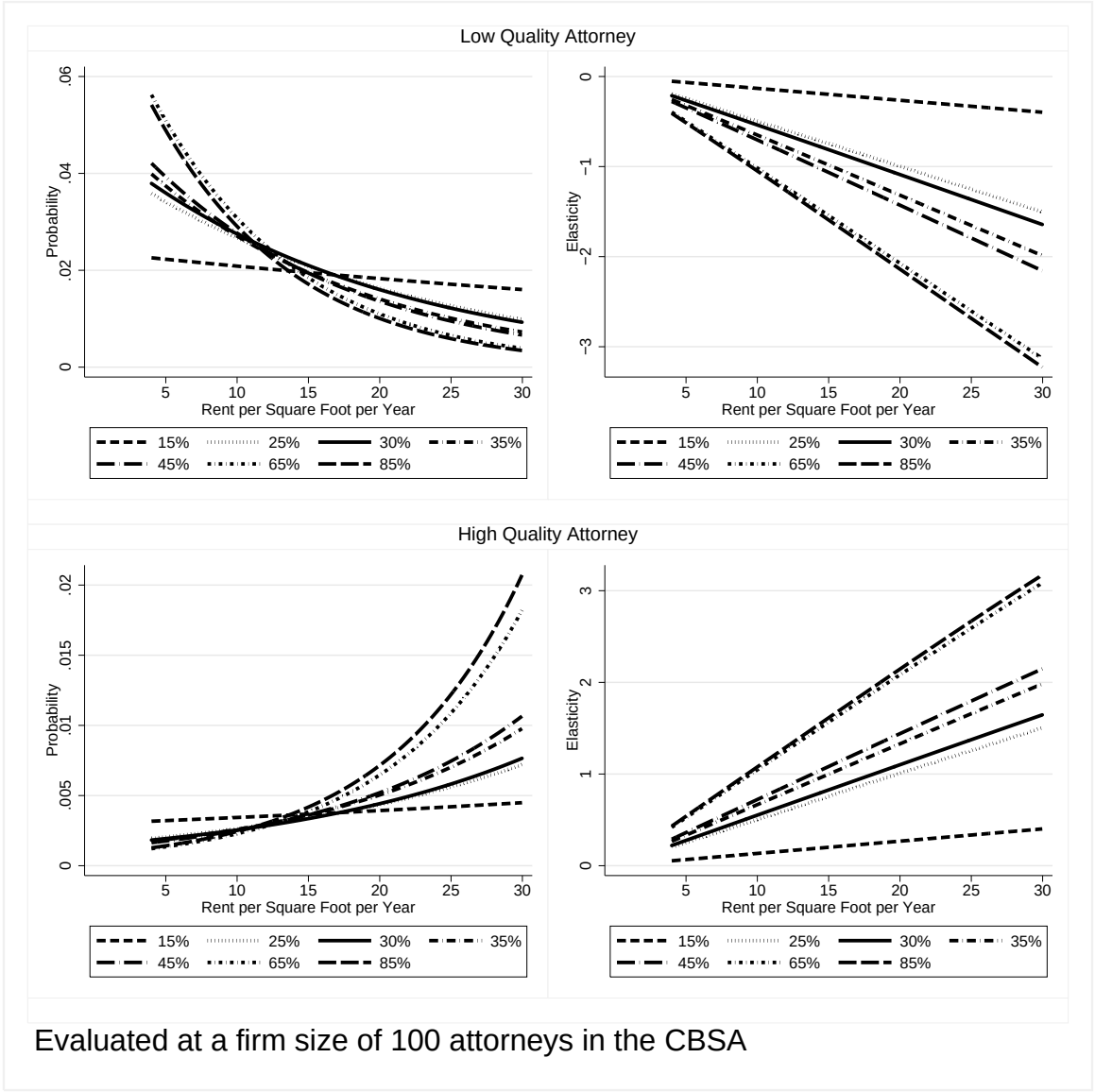
Evaluated at a firm size of 1 attorney in the CBSA

Figure C.5: Probabilities and elasticities using predicted rent and alternative bandwidths: model 3



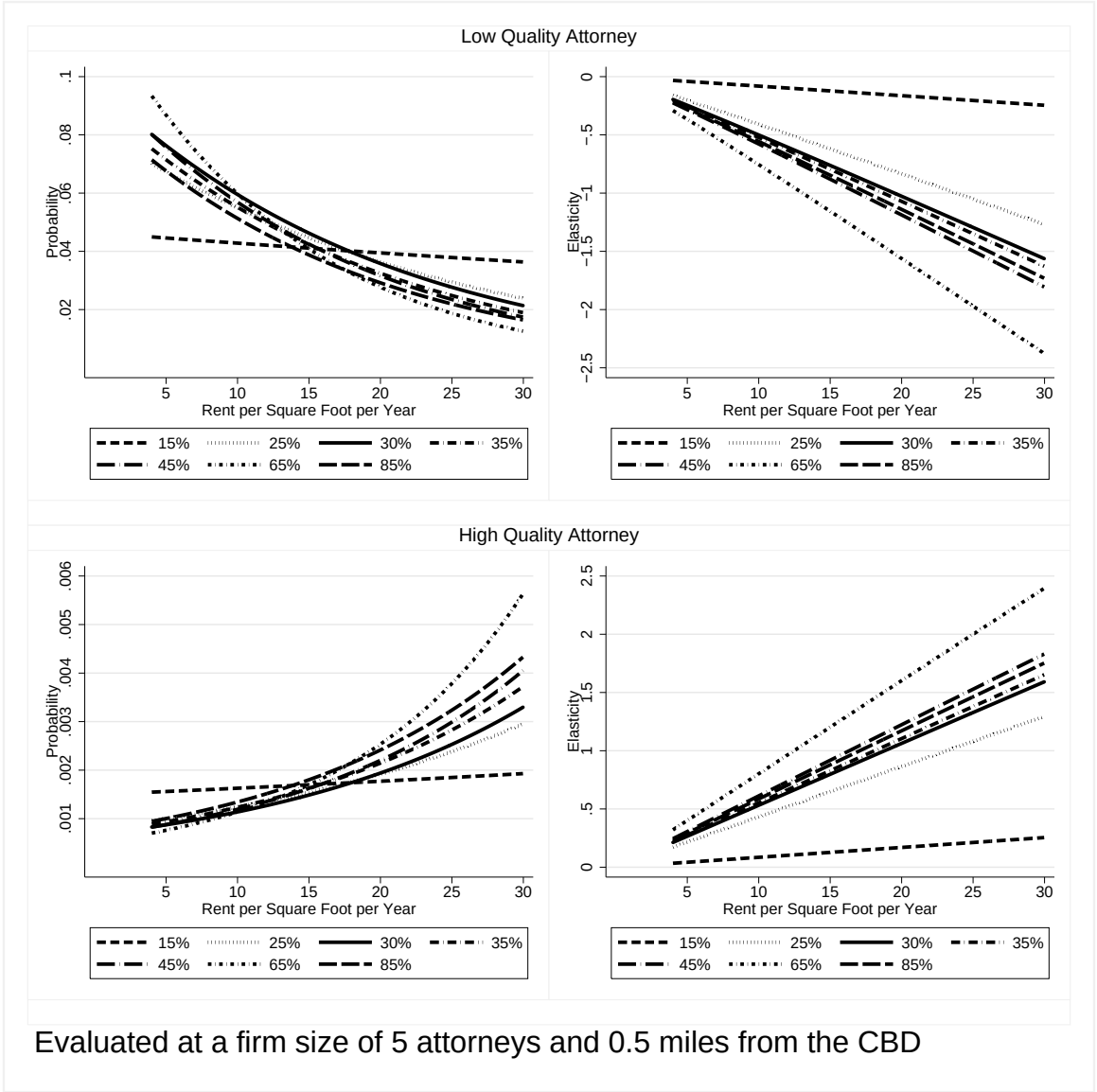
Evaluated at a firm size of 5 attorneys in the CBSA

Figure C.6: Probabilities and elasticities using predicted rent and alternative bandwidths: model 3



Evaluated at a firm size of 100 attorneys in the CBSA

Figure C.7: Probabilities and elasticities using predicted rent and alternative bandwidths: model 4



Tables

Table 3.1: Private sector employers with more than 50 attorneys in the Cleveland CBSA

Employer	Attorneys
jones day	255
baker & hostetler	181
calfee, halter & griswold llp	151
thompson hine llp	151
squire, sanders & dempsey	149
ulmer & berne llp	118
benesch friedlander coplan & aronoff	106
tucker ellis & west llp	91
hahn loeser & parks llp	83
mcdonald hopkins llc	82
progressive insurance	79
key bank national association	68
reminger co., l.p.a.	66
frantz ward llp	63
taft, stettinius & hollister llp	63
national city bank	55
gallagher sharp	51
roetzel & andress	51

Table 3.2: Highest ranked U.S. law schools

US News Rank (2009)	Law School	Median Private Sector Starting Salary	Private Sector Attorneys in the Cleveland CBSA
1	Yale U.	160,000	23
2	Harvard U.	160,000	109
3	Stanford U.	160,000	9
4	Columbia U.	160,000	26
5	New York U.	160,000	25
6	U. of California-Berkeley	160,000	4
6	U. of Chicago	160,000	28
8	U. of Pennsylvania	160,000	37
9	U. of Michigan-Ann Arbor	160,000	152
10	Duke U.	160,000	45
10	Northwestern U.	160,000	24
10	U. of Virginia	160,000	59

Table 3.3: Most common law schools attended by Cleveland private sector attorneys

Law School	Location	Private Sector Attorneys in the Cleveland CBSA	Rank of Law School	Median Private Sector Starting Salary
Cleveland State U.	Cleveland, OH	3363	Tier 3	80,000
Case Western Reserve U.	Cleveland, OH	1855	55	110,000
Ohio State U.	Columbus, OH	499	35	100,000
U. of Akron	Akron, OH	499	Tier 3	63,000
U. of Toledo	Toledo, OH	192	Tier 3	70,000
U. of Michigan-Ann Arbor	Ann Arbor, MI	152	9	160,000
Harvard U.	Cambridge, MA	109	2	160,000
Ohio Northern U.	Ada, OH	105	Tier 3	60,000
Capital U.	Columbus, OH	98	Tier 4	79,000
Georgetown U.	Washington, DC	84	14	160,000

Table 3.4: Spatial distribution of private sector attorneys in the Cleveland CBSA

Distance from the CBD	Attorneys	Attorneys per Square Mile	Inexperienced * Attorneys	Inexperienced * Attorneys per Square Mile
less than 0.1 mile	1397	45064.5	259	8354.84
0.1 to 0.25 mile	686	4157.6	110	666.67
0.25 to 0.5 mile	2029	3444.8	357	606.11
0.5 to 1 mile	229	113.3	28	13.85
1 to 5 miles	253	5.6	30	0.66
5 to 10 miles	1104	7.8	102	0.72
10 to 15 miles	1458	6.4	150	0.66
15 to 20 miles	440	1.6	38	0.14
20 to 25 miles	284	0.9	24	0.08
more than 25 miles	403	0.4	33	0.03
all distances	8283	4.1	1131	0.56

* - Inexperienced attorneys are attorneys that have been members of the Ohio Bar for less than 6 years

Table 3.5: Descriptive statistics of private sector attorneys in the Cleveland CBSA

	Minimum	1st Quartile	Median	3rd Quartile	Maximum	Mean	Std Dev
Age	25.0	39.0	49.0	58.0	96.0	49.4	12.9
Experience	0.0	10.0	20.0	30.0	70.0	20.7	13.0
Male	0.0	0.0	1.0	1.0	1.0	0.7	0.4
Miles To CBD	0.0	0.2	0.5	11.4	41.0	6.4	8.1
Firm Size in CBSA	1.0	1.0	5.0	36.0	255.0	35.5	60.3

Table 3.6: Information provided in CoStar property database

Variable	Values
PropertyID	Unique CoStar identifier
Building Address	
City	
Rentable Building Area	sq. ft.
Vacancy Rate	
Building Class	A, B, C
Year Built	
Year Renovated	
Number Of Stories	
LEED Building	Yes or No
EnergyStar Building	Yes or No
Banking	Yes or No
Convenience Store	Yes or No
Dry Cleaner	Yes or No
Fitness Center	Yes or No
Food Court	Yes or No
Food Service	Yes or No
Mail Room	Yes or No
Restaurant	Yes or No

Table 3.7: Information provided in CoStar space database

Variable	Values
PropertyID	Unique CoStar identifier
SuiteID	
Floor	Numeric value, Basement, Ground, Lobby, Lower Level, Mezzanine, Penthouse
Lease Type	New, Relet, or Sublet
Space Use	Flex, Medical, Office, Office/Medical, Office/Retail, Retail, or Warehouse
SF For Lease	sq. ft.
Asking Rent	dollars per square foot per year
Services in Rent	Plus Elec and Clean, Double Net, Full Service Gross, Modified Gross, Negotiable, Net, Plus All Utilities, Plus Cleaning, Plus Electric, TBD, Tenant Electric, Triple Net, or Utilities and Char

Table 3.8: Descriptive statistics of CoStar property data in the Cleveland CBSA

	Minimum		Maximum		Mean		Standard Deviation	
	All	In Sample	All	In Sample	All	In Sample	All	In Sample
Amenities	0.0	0.0	1	1	0.07	0.20	0.26	0.40
Building Age	-2.0	0.0	208	162	51.00	43.33	34.59	33.91
Building Size (,000 sq ft)	0.3	0.6	1463	1323	27.91	57.15	79.41	132.57
Class A	0.0	0.0	1	1	0.03	0.09	0.17	0.29
Class B	0.0	0.0	1	1	0.36	0.52	0.48	0.50
Class C	0.0	0.0	1	1	0.61	0.39	0.49	0.49
Miles To CBD	0.0	0.0	40	39	13.91	12.27	9.01	8.11
Number of Floors	1.0	1.0	57	57	2.20	3.35	2.73	4.73
Renovated	0.0	0.0	1	1	0.12	0.25	0.32	0.43
Vacancy Rate	0.0	0.0	1	1	0.12	0.29	0.28	0.30

Table 3.9: Descriptive statistics of CoStar space data in the Cleveland CBSA

	Min	Max	Mean	Std Dev
Building Age	0.0	162	47.37	31.62
Amenities	0.0	1	0.49	0.50
Class A	0.0	1	0.14	0.35
Class B	0.0	1	0.63	0.48
Class C	0.0	1	0.23	0.42
Miles To CBD	0.0	39	8.77	7.73
Number of Floors	1.0	57	7.90	9.35
Renovated	0.0	1	0.34	0.48
Building Size (,000 sq ft)	0.6	1323	157.38	238.23
Vacancy Rate	0.0	1	0.31	0.26
Asking Rent	1.1	45	14.72	4.29
Full Gross Rent	0.0	1	0.43	0.49
Modified Gross Rent	0.0	1	0.46	0.50
Net Rent	0.0	1	0.11	0.32
Sq Feet of Space	55.0	95000	4333.48	7243.36

Table 3.10: Binary logit estimate of the difference between properties with space for lease and those without space for lease

Dependent Variable: Space for Lease = 1	
Class A	0.750
	(0.282)**
Class B	0.339
	(0.115)**
ln of Rentable Building Area (sq. ft.)	0.422
	(0.047)**
Building Height (Floors)	-0.013
	(0.017)
Fraction of Building Occupied	-2.626
	(0.176)**
ln of Building Age	-0.285
	(0.070)**
Amenities	0.791
	(0.170)**
Renovated	0.966
	(0.130)**
Pseudo R^2	0.22
Observations	3,436
χ^2	558.77
Prob $> \chi^2$	0.00

* $p < 0.05$; ** $p < 0.01$

Table 3.11: Least squares estimate of the rent gradient in the Cleveland CBSA

Dependent Variable: ln of Asking Rent	
ln of Miles to CBD	-0.010
	(0.005)*
Net Rent	-0.194
	(0.019)**
Modified Gross Rent	-0.084
	(0.011)**
ln of Size of Space (sq. ft.)	-0.014
	(0.005)**
ln of Rentable Building Area (sq. ft.)	0.022
	(0.010)*
Class A	0.330
	(0.023)**
Class B	0.181
	(0.017)**
ln of Building Age	-0.155
	(0.012)**
Amenities	0.024
	(0.012)
Renovated	0.043
	(0.014)**
Fraction of Building Occupied	0.141
	(0.024)**
ln of Building Height (Floors)	0.021
	(0.012)
R^2	0.43
Observations	2,840
F(12 , 2827)	181.81
Prob > F	0.0000

* $p < 0.05$; ** $p < 0.01$

Table 3.12: Ordered logit estimate of the attorney quality gradient from the distance model

Miles To CBD	-0.043
	(0.008)**
Pseudo R^2	0.01
Observations	8,283
χ^2	28.22
Prob $> \chi^2$	0.00

* $p < 0.05$; ** $p < 0.01$

Table 3.13: Binary logit estimate of the difference between matched and non-matched properties

Dependent Variable: Successful Match = 1	
Attorney Quality	0.003
	(0.002)*
Miles To CBD	-0.134
	(0.027)**
Attorneys of Firm in the CBSA	-0.006
	(0.005)
Male	0.079
	(0.081)
Age of Attorney	-0.033
	(0.008)**
Experience	0.018
	(0.008)*
Pseudo R^2	0.14
Observations	8,311
χ^2	53.14
Prob $> \chi^2$	0.00

* $p < 0.05$; ** $p < 0.01$

Table 3.14: Ordered logit estimates of the attorney quality gradient using matched properties

	Model 1	Model 2	Model 3	Model 4
Rent	0.094	0.090	0.039	0.039
	(0.008)**	(0.009)**	(0.009)**	(0.009)**
Miles To CBD		-0.019		-0.001
		(0.010)		(0.008)
Attorneys of Firm in the CBSA			0.009	0.009
			(0.001)**	(0.001)**
Pseudo R^2	0.02	0.02	0.03	0.03
Observations	4,100	4,100	4,100	4,100
χ^2	142.98	151.08	275.69	278.03
Prob $> \chi^2$	0.00	0.00	0.00	0.00

* $p < 0.05$; ** $p < 0.01$

Table 3.15: Explanatory variables of the locally weighted regressions

Variable	Value for Prediction
Full Gross Rent (excluded)	0
Modified Gross Rent	1
Net Rent	0
Class A Building	0
Class B Building	1
Class C Building (excluded)	0
ln of Building Age (plus one)	ln(37+1)
Building has Amenities	0
Building Renovated	0
Fraction of Building Occupied	0.7637
ln of Rentable Building Area (sq. feet)	ln(61516)
ln of Size of Space (sq. feet)	ln(1975)
9 knot linear spline of distance to the CBD	based on attorney location

Table 3.16: Summary statistics from the locally weighted regressions

Diagnostic	Minimum	1st Quartile	Median	3rd Quartile	Maximum	Mean	Std Dev
Predicted Rent	5.76	13.21	14.41	15.22	24.38	14.34	1.93
R^2	0.35	0.47	0.52	0.63	0.78	0.54	0.09
Minimum Distance in Bandwidth	0.00	0.05	0.32	0.80	10.44	0.57	0.82
Median Distance in Bandwidth	0.23	5.03	6.78	10.34	31.32	8.36	5.65
Maximum Distance in Bandwidth	0.99	6.88	10.24	14.31	38.52	11.35	6.91
Number of Observations	790.00	847.00	853.00	855.00	898.00	850.52	11.22

Table 3.17: Hypothesis tests from the locally weighted regressions

Variable	Median Coefficient	Negative	Zero	Positive
Net Rent	-2.37	85.4%	14.6%	0.0%
Modified Gross Rent	-1.47	65.8%	28.3%	5.9%
ln of Size of Space (sq. ft.)	-0.30	37.3%	54.3%	8.4%
ln of Rentable Building Area (sq. ft.)	0.41	0.0%	63.5%	36.5%
Class A	4.85	0.0%	10.5%	89.5%
Class B	1.83	0.0%	19.1%	80.9%
ln of Building Age	-1.59	95.3%	4.7%	0.0%
Amenities	0.36	21.8%	42.4%	35.8%
Renovated	0.15	13.3%	80.7%	6.0%
Fraction of Building Occupied	1.14	0.4%	69.4%	30.2%

Table 3.18: Ordered logit estimates of the attorney quality gradient using predicted rent

	Model 1	Model 2	Model 3	Model 4
Rent	0.119	0.104	0.055	0.053
	(0.019)**	(0.030)**	(0.011)**	(0.015)**
Miles To CBD		-0.009		-0.001
		(0.011)		(0.005)
Attorneys of Firm in the CBSA			0.009	0.009
			(0.001)**	(0.001)**
Pseudo R^2	0.01	0.01	0.03	0.03
Observations	8,283	8,283	8,283	8,283
χ^2	60.04	53.64	330.70	348.78
Prob $> \chi^2$	0.00	0.00	0.00	0.00

* $p < 0.05$; ** $p < 0.01$

Table 3.19: Ordered logit estimates of the attorney quality gradient using predicted rent: effect of experience

	Full Sample	Experience < 6	Experience ≥ 6	Experience ≥ 20
Rent	0.053	0.034	0.061	0.067
	(0.015)**	(0.025)	(0.016)**	(0.021)**
Miles To CBD	-0.001	-0.006	-0.000	0.003
	(0.005)	(0.014)	(0.006)	(0.007)
Attorneys of Firm in the CBSA	0.009	0.010	0.009	0.010
	(0.001)**	(0.002)**	(0.001)**	(0.001)**
Pseudo R^2	0.03	0.05	0.03	0.04
Observations	8,283	1,131	7,152	4,157
χ^2	348.78	64.97	264.74	204.80
Prob $> \chi^2$	0.00	0.00	0.00	0.00

* $p < 0.05$; ** $p < 0.01$

Table 3.20: Ordered logit estimates of the attorney quality gradient using predicted rent: effect of school location

	Full Sample	In State School	Out of State School
Rent	0.053	0.031	0.080
	(0.015)**	(0.017)	(0.018)**
Miles To CBD	-0.001	-0.009	0.008
	(0.005)	(0.006)	(0.011)
Attorneys of Firm in the CBSA	0.009	0.007	0.004
	(0.001)**	(0.001)**	(0.001)**
Pseudo R^2	0.03	0.03	0.01
Observations	8,283	6,744	1,539
χ^2	348.78	104.49	102.62
Prob $> \chi^2$	0.00	0.00	0.00

* $p < 0.05$; ** $p < 0.01$

Table 4.1: Naive wage gradients in the Cleveland CBSA

	Log-linear (LS)	Log-linear (Poisson)	Semi-log (Poisson)	Linear (LS)
ln of Miles To CBD (+1)	-0.040	-0.043		
	(0.008)**	(0.009)**		
Miles To CBD			-0.006	-0.587
			(0.001)**	(0.113)**
Observations	8,247	8,247	8,247	8,247
R^2	0.03			0.03
Pseudo R^2		0.02	0.02	

* $p < 0.05$; ** $p < 0.01$

Table 4.2: Estimated wage gradients in the Cleveland CBSA

Miles to CBD	Linear Naive	Semi-log Naive	Log-linear Naive	Fixed Rent (Model 2)	Fixed Rent (Model 4)
0	99,960	100,033	101,697	97,376	92,916
1	99,373	99,406	98,682	97,265	92,902
2	98,787	98,782	96,959	97,155	92,889
3	98,200	98,163	95,756	97,045	92,876
4	97,614	97,547	94,832	96,935	92,863
5	97,027	96,935	94,084	96,826	92,849
6	96,441	96,327	93,457	96,717	92,836
7	95,854	95,723	92,916	96,608	92,823
8	95,268	95,123	92,442	96,499	92,810
9	94,681	94,527	92,020	96,391	92,796
10	94,095	93,934	91,640	96,283	92,783
15	91,162	91,025	90,161	95,748	92,717
20	88,230	88,206	89,103	95,221	92,651
25	85,297	85,475	88,281	94,702	92,586
30	82,364	82,828	87,609	94,190	92,520
35	79,432	80,264	87,042	93,687	92,455
40	76,499	77,778	86,552	93,191	92,389
45	73,567	75,370	86,120	92,704	92,324
50	70,634	73,036	85,735	92,225	92,259

Table B.1: Ordered logit estimates of the attorney quality gradient under alternative error structures: model 1

	GWR-IID	GWR-nonIID-1	GWR-nonIID-2	LS-nonIID
Rent	0.119	0.119	0.119	0.122
	(0.019)**	(0.026)**	(0.017)**	(0.016)**
Pseudo R^2	0.01	0.01	0.01	0.01
Observations	8,283	8,283	8,283	8,283
χ^2	60.04	60.04	60.04	66.11
Prob $> \chi^2$	0.00	0.00	0.00	0.00

* $p < 0.05$; ** $p < 0.01$

Table B.2: Ordered logit estimates of the attorney quality gradient under alternative error structures: model 2

	GWR-IID	GWR-nonIID-1	GWR-nonIID-2	LS-nonIID
Rent	0.104	0.104	0.104	0.104
	(0.030)**	(0.037)**	(0.028)**	(0.032)**
Miles to CBD	-0.009	-0.009	-0.009	-0.009
	(0.011)	(0.013)	(0.011)	(0.013)
Pseudo R^2	0.01	0.01	0.01	0.01
Observations	8,283	8,283	8,283	8,283
χ^2	53.64	53.64	53.64	57.00
Prob $> \chi^2$	0.00	0.00	0.00	0.00

* $p < 0.05$; ** $p < 0.01$

Table B.3: Ordered logit estimates of the attorney quality gradient under alternative error structures: model 3

	GWR-IID	GWR-nonIID-1	GWR-nonIID-2	LS-nonIID
Rent	0.055	0.055	0.055	0.056
	(0.011)**	(0.011)**	(0.011)**	(0.011)**
Attorneys of Firm in the CBSA	0.009	0.009	0.009	0.009
	(0.001)**	(0.001)**	(0.001)**	(0.001)**
Pseudo R^2	0.03	0.03	0.03	0.03
Observations	8,283	8,283	8,283	8,283
χ^2	330.70	330.70	330.70	362.16
Prob $> \chi^2$	0.00	0.00	0.00	0.00

* $p < 0.05$; ** $p < 0.01$

Table B.4: Ordered logit estimates of the attorney quality gradient under alternative error structures: model 4

	GWR-IID	GWR-nonIID-1	GWR-nonIID-2	LS-nonIID
Rent	0.053	0.053	0.053	0.053
	(0.015)**	(0.015)**	(0.015)**	(0.015)**
Miles to CBD	-0.001	-0.001	-0.001	-0.002
	(0.005)	(0.005)	(0.005)	(0.006)
Attorneys of Firm in the CBSA	0.009	0.009	0.009	0.009
	(0.001)**	(0.001)**	(0.001)**	(0.001)**
Pseudo R^2	0.03	0.03	0.03	0.03
Observations	8,283	8,283	8,283	8,283
χ^2	348.78	348.78	348.78	398.74
Prob $> \chi^2$	0.00	0.00	0.00	0.00

* $p < 0.05$; ** $p < 0.01$

Table C.1: Ordered logit estimates of the attorney quality gradient under alternative bandwidths: model 1

	15%	25%	30%	35%	45%	65%	85%
Rent	0.034	0.117	0.119	0.140	0.149	0.225	0.218
	(0.011)**	(0.019)**	(0.019)**	(0.016)**	(0.018)**	(0.073)**	(0.114)
Pseudo R^2	0.01	0.01	0.01	0.01	0.01	0.01	0.00
Observations	8,283	8,283	8,283	8,283	8,283	8,283	8,283
χ^2	11.99	63.24	60.04	99.51	95.44	20.05	6.91
Prob $> \chi^2$	0.00	0.00	0.00	0.00	0.00	0.00	0.01

* $p < 0.05$; ** $p < 0.01$

Table C.2: Ordered logit estimates of the attorney quality gradient under alternative bandwidths: model 2

	15%	25%	30%	35%	45%	65%	85%
Rent	0.020	0.102	0.104	0.098	0.103	0.134	0.057
	(0.007)**	(0.026)**	(0.030)**	(0.033)**	(0.040)**	(0.085)	(0.113)
Miles to CBD	-0.030	-0.009	-0.009	-0.022	-0.022	-0.025	-0.038
	(0.008)**	(0.010)	(0.011)	(0.011)*	(0.012)	(0.010)**	(0.009)**
Pseudo R^2	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Observations	8,283	8,283	8,283	8,283	8,283	8,283	8,283
χ^2	37.69	54.16	53.64	63.93	62.96	31.43	30.75
Prob $> \chi^2$	0.00	0.00	0.00	0.00	0.00	0.00	0.00

* $p < 0.05$; ** $p < 0.01$

Table C.3: Ordered logit estimates of the attorney quality gradient under alternative bandwidths: model 3

	15%	25%	30%	35%	45%	65%	85%
Rent	0.013	0.051	0.055	0.067	0.072	0.105	0.108
	(0.004)**	(0.011)**	(0.011)**	(0.014)**	(0.015)**	(0.029)**	(0.037)**
Attorneys of Firm in the CBSA	0.009	0.009	0.009	0.009	0.009	0.009	0.010
	(0.001)**	(0.001)**	(0.001)**	(0.001)**	(0.001)**	(0.001)**	(0.001)**
Pseudo R^2	0.03	0.03	0.03	0.03	0.03	0.03	0.03
Observations	8,283	8,283	8,283	8,283	8,283	8,283	8,283
χ^2	144.06	293.06	330.70	298.35	259.38	156.02	108.38
Prob $> \chi^2$	0.00	0.00	0.00	0.00	0.00	0.00	0.00

* $p < 0.05$; ** $p < 0.01$

Table C.4: Ordered logit estimates of the attorney quality gradient under alternative bandwidths: model 4

	15%	25%	30%	35%	45%	65%	85%
Rent	0.009	0.043	0.053	0.055	0.061	0.080	0.059
	(0.003)*	(0.016)**	(0.015)**	(0.014)**	(0.016)**	(0.035)*	(0.046)
Miles to CBD	-0.013	-0.004	-0.001	-0.007	-0.006	-0.007	-0.013
	(0.004)**	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)*
Attorneys of Firm in the CBSA	0.009	0.009	0.009	0.009	0.009	0.009	0.009
	(0.001)**	(0.001)**	(0.001)**	(0.001)**	(0.001)**	(0.001)**	(0.001)**
Pseudo R^2	0.03	0.03	0.03	0.03	0.03	0.03	0.03
Observations	8,283	8,283	8,283	8,283	8,283	8,283	8,283
χ^2	185.09	272.61	348.78	427.75	370.28	154.37	131.39
Prob $> \chi^2$	0.00	0.00	0.00	0.00	0.00	0.00	0.00

* $p < 0.05$; ** $p < 0.01$